



Published in final edited form as:

Dev Cell. 2018 September 10; 46(5): 651–667.e10. doi:10.1016/j.devcel.2018.07.025.

A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-seq

Christopher Daniel Green^{*,1}, Qianyi Ma^{*,1}, Gabriel L. Manske², Adrienne Niederriter Shami¹, Xianing Zheng¹, Simone Marini⁴, Lindsay Moritz², Caleb Sultan¹, Stephen J. Gurczynski³, Bethany B. Moore^{3,5}, Michelle D. Tallquist⁶, Jun Z. Li^{†,1,4}, Saher Sue Hammou^{†,1,2,7,8,9}

¹Department of Human Genetics, University of Michigan, Ann Arbor, MI

²Cellular and Molecular Biology Program, University of Michigan, Ann Arbor, MI

³Department of Internal Medicine, University of Michigan, Ann Arbor, MI

⁴Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI

⁵Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI

⁶University of Hawaii Cancer Center, Honolulu, HI

⁷Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, MI

⁸Department of Urology, University of Michigan, Ann Arbor, MI

⁹Lead Contact

Summary:

Spermatogenesis requires intricate interactions between the germline and somatic cells. Within a given cross-section of a seminiferous tubule, multiple germ and somatic cell types co-occur. This cellular heterogeneity has made it difficult to profile distinct cell types at different stages of development. To address this challenge, we collected single-cell RNA sequencing data from ~35K cells from the adult mouse testis, and identified all known germ and somatic cells, as well as two unexpected somatic cell types. Our analysis revealed a continuous developmental trajectory of germ cells from spermatogonia to spermatids, and identified novel candidate transcriptional regulators at several transition points during differentiation. Focused analyses delineated four subtypes of spermatogonia and nine subtypes of Sertoli cells, the latter linked to histologically defined developmental stages over the seminiferous epithelial cycle. Overall, this high-resolution cellular atlas represents a community resource and foundation of knowledge to study germ cell development and *in vivo* gametogenesis.

Correspondence should be addressed to: hammou@med.umich.edu or junzli@med.umich.edu.

*Denotes Equal First Authors

†Denotes co-corresponding authors

Author contributions

S.S.H. overall project design. J.Z.L. oversaw the computational analysis. C.D.G, G.L.M., A.N.S., C.S., B.B.M., L.E.M, and S.J.G. performed experiments. Q.M., and J.Z.L analyzed data. S.M. compared our data with Mouse Cell Atlas paper. S.S.H. wrote the manuscript with help from J.Z.L., C.D.G., and Q.M. Comments from all authors were provided.

Declaration of interests

The authors have no competing interests.

Keywords

Spermatogenesis; spermatogonial stem cell; heterogeneity; single-cell RNA-seq; germ cell developmental trajectory; testis niche

Introduction

Spermatogenesis is a complex process by which diploid spermatogonial stem cells terminally differentiate to produce mature, haploid spermatozoa within the testis. This process is continuous throughout adult life in mammals, and is characterized by three phases: mitotic proliferation, meiosis, and spermiogenesis (a period of morphological maturation and chromatin repackaging). Over the past decade, considerable efforts have been devoted to examining the germ cell intrinsic programs by isolating and analyzing specific germ cell populations using a variety of molecular or genetic approaches (Chan et al., 2014; Costoya et al., 2004; DeFalco et al., 2015; Evans et al., 2014; Guo et al., 2014; Hammoud et al., 2014; Hammoud et al., 2015; Hara et al., 2014; Hermann et al., 2015; Inoue et al., 2017; Johnston et al., 2008; Kliesch et al., 1992; Lesch et al., 2013; Oatley et al., 2006; Parvinen et al., 1992; Soderstrom and Parvinen, 1976; Zimmermann et al., 2015). Although these methodologies have provided valuable insights, our understanding of germ cell differentiation is limited because the analyses are restricted to selected subsets of germ cell populations that can be isolated via cell surface markers or transgenic lines.

Successful execution of the germ cell developmental program requires ongoing juxtacrine, paracrine, and endocrine signaling between germ cells, supporting somatic cells, and the pituitary gland (Chen et al., 2016; DeFalco et al., 2015; Eddy, 2002; Franca et al., 1998; Griswold, 1995; Jegou, 1993; Meng et al., 2000; O'Shaughnessy et al., 2008; Phillips et al., 2010; Sharpe, 1986). Within the testis there are multiple somatic cell types that produce growth factors that influence neighboring somatic cells or germ cell development, either by direct contact or by indirect, ligand-mediated signaling (DeFalco et al., 2015; Hofmann et al., 2005; Kubota et al., 2004; Maekawa et al., 1996; Meng et al., 2000; Moore and Morris, 1993; Nalbandian et al., 2003; Oatley et al., 2009; Smith and Walker, 2014). These include: Sertoli cells (secreted factors such as *Gdnf*, *Fgf*, *Egf*), macrophages (*Csf1*), endothelial cells (*Vegf*), steroidogenic Leydig cells (testosterone, *Igf1*, *Csf1*), and peritubular myoid cells. Although the somatic cells of the testis provide essential support to the germline throughout spermatogenesis, our understanding of their molecular subtypes, regulatory programs, and germline-soma or soma-soma communications remain poorly understood. Sequencing-based profiling of single-cell transcriptomes now provide a cost-effective method to survey thousands of cells to define functional subtypes and their molecular signatures. Such information naturally leads to the identification of both known and previously unknown cell types, including transient populations that were too rare to be detected with low-throughput approaches. Importantly, the unbiased characterization of functional heterogeneity at the single-cell level and the associated marker genes will be essential for understanding inter-cellular interactions in the native organ structure, and for probing the spatiotemporal patterns of signaling among germ cells and supporting cells. Such a resource is expected to improve

the knowledge base for future studies of germ cell biology, and ultimately improve targeted therapies for male infertility.

In this study we used the Drop-seq technology (Macosko et al., 2015) to analyze 34,633 single cells isolated from the mouse testis, consisting of both unselected cells and targeted enrichment of rare cell types. This large number of cells allowed us to generate a detailed cellular and molecular atlas that includes not only known cell types in the testis, but also two previously undescribed adult somatic cell populations. Our study reveals the continuous nature of germ cell development, identifies rare and key development transitions, and uncovers known and novel candidate transcriptional regulators that accompany germ cell differentiation. Furthermore, iterative re-clustering of major cell populations reveals a deeper level of heterogeneity contained within the spermatogonia and Sertoli cells, and single molecule RNA detection methods, including single-molecule fluorescence *in situ* hybridization (smFISH) and single-molecule hybridization chain reaction (smHCR), allowed us to spatially map cellular subtypes to histological stages described a few decades ago.

Results

Single cell RNA-seq establishes a detailed atlas of cellular heterogeneity in the adult mouse testis.

To provide an unbiased survey of cellular diversity of the adult mouse testis, we applied the Drop-seq technology to characterize single-cell transcriptomes, initially in six replicate experiments of the whole testis. Each replicate captures ~2,000 cells from a C57BL/6J male, ranging in age from 7- to 9-weeks old. From our human-mouse mixing experiment, consisting of 25% spike-in of human A549 cells and 75% mouse testicular cells, we determined that the human-mouse doublet rates were <1.8% (Figure S1A), thus confirming that key experimental parameters such as cell concentration and flow rates have been optimized for capturing and analyzing single cells. The six independent experiments revealed similar patterns of cellular heterogeneity (Figure S1B), and consistent clustering solutions (Figure S1C). The concordance among the six datasets allowed us to merge datasets to create a combined collection of ~12,000 cells from the adult mouse testis.

Using previously described cell-type specific markers we identified all major germ cell populations covering the full developmental spectrum, including spermatogonia (SPG), meiotic spermatocytes (SCytes), postmeiotic haploid round spermatids (STids) and elongating spermatids (ES). However, the gonadal somatic cell compartment was underrepresented in the total testis datasets. To increase the representation of these rarer somatic cells we generated another 19 datasets by 1) applying gentler dissociation methods, 2) enriching for interstitial cells by depleting germ cells, 3) using mouse transgenic lines to enrich specifically for spermatogonia and Sertoli cells (e.g. *Amh-cre;mTmG*, *Sox9-eGFP*, *Gfra1-creERT2;mTmG*), or 4) enriching for spermatogonia, immune, Leydig or interstitial cells using a series of cell surface markers (e.g. Thy1, Kit, Sca1) (see Methods, Figure S1D, Table S1 – summarizes the datasets generated, the number of cells per dataset, and number of detected transcripts or genes per cell). Therefore, as a result of our 25 Drop-Seq experiments, we have analyzed approximately 35K cells (post QC filters). Systematic assessment of experimental batch effects confirmed that the identification of major cell types

was not driven by batch-batch variation, as shown by reliable alignment of identified cell types across batches (see Methods). Each cell has an average number of 6,205 UMIs, and 2,057 genes detected. This sequencing depth and the number of detected genes are comparable to previously published reports using the Drop-seq technology, and is sufficient for defining distinct cell types (see methods) (Campbell et al., 2017; Heimberg et al., 2016; Macosko et al., 2015; Stephenson et al., 2018; Tanay and Regev, 2017)

Unsupervised clustering of the ~35K cells identified 11 major cell types. Expression patterns of known marker genes and gene ontology analysis assigned the 11 cell types to the four known germ cell populations described above, as well as seven somatic cell populations (Figure 1A, B, S1E). The seven somatic cell populations included five known somatic cell populations: Sertoli, myoid, Leydig, endothelial, and macrophages (Figure 1B, C, Table S2), and two unexpected cell populations: an innate lymphoid (type II) cell population (ILC II) and a novel mesenchymal cell population (See the somatic sections below). Taken together, our collection of ~35K single cells illustrates the functional diversity among testis cells and identifies known and novel major cell types. The enrichment experiments, in particular, provided us a unique opportunity to focus on major cell types individually and, as we show below, to delineate previously unappreciated subtypes. This step-wise exploration of functional subtypes also defines the genes and pathways underlying their biological differences at increasing levels of granularity (Figure 1A).

Extracting functional properties of individual cells uncovers between- and within-cell type heterogeneity.

To correlate the transcriptomic properties of single cells with previously described cytological features or average attributes of bulk cell populations, we calculated several per-cell transcriptome-derived attributes, including the percentages of transcripts accounted for by chromosomes X and Y genes, by mitochondria-encoded RNAs, the total number of detected genes per cell, the total number of unique transcripts, and the Gini index of each cell (**Table uploaded to GEO: GSE112393**). These single-cell indices represent an important component of our cellular atlas, allowing extensive comparisons with the existing knowledge of the basic biology of germ cells and their somatic supporting cells. Below we highlight four illustrative examples.

First, when examining the proportion of ChrX transcripts in the entire transcriptome, we find that this value is highest in Sertoli cells, followed by other somatic and spermatogonia cell populations, is greatly reduced in spermatocytes, and partially recovers in spermatids (Figure 1D). Similarly, the proportion of ChrY transcripts is consistent across most cell types, becomes depleted in spermatocytes, and recovers in spermatids. The extremely low levels of X and Y transcripts in spermatocytes are consistent with the timing of XY body formation during meiosis, a specialized nuclear territory for ChrX and ChrY where both transcription and homologous recombination are suppressed (Handel, 2004; Hoyer-Fender, 2003; McKee and Handel, 1993; Solari, 1974). After meiotic sex chromosome inactivation, both chromosomes are reactivated in post-meiotic cells (Mueller et al., 2008). Interestingly, our results show that, while Y-chromosome transcripts are transiently elevated only in STids, X-chromosome transcripts are present in both STids and ES, suggesting that either the X-

chromosome transcripts persist longer than ChrY transcripts, or maintain a longer transcriptional activity.

Second, it has been known that syncytial development of germ cells is an evolutionarily conserved process from fruit flies to mammals, in both the male and the female germlines (Pepling et al., 1999). These intercellular bridges presumably allow *de facto* sharing of cytoplasmic content, and ensure synchronous development and gametic equivalence of a set of clonally related germ cells (Braun et al., 1989; Fawcett et al., 1959). Evidence for cytoplasmic sharing in connected haploid round spermatids was first demonstrated using a hemizygous transgenic mouse that expresses the human growth hormone transgene under the control of a round spermatid specific Protamine 1 promoter (Braun et al., 1989). In that study, ~90% of the round and elongating spermatids express human growth hormone protein despite a prediction that 50% of the haploid cells in the hemizygous animal would express the transgene. To discern if our data also show evidence of cytoplasmic sharing, we examined the distribution of X and Y transcripts in the 9,923 haploid round spermatids, which should bear either an X or a Y chromosome. Remarkably similar to the previous findings, ~86% of the round spermatids (N=8,531) contain both X and Y transcripts (Figure S1F). Since steady-state levels of mRNA may persist long after the initial transcription, and the fraction of detection depends on the depth of sequencing of individual cells, we contrasted the levels of X and Y transcripts in diploid SPG cells to haploid round spermatids, separately comparing groups of cells with a comparable range of total UMI. For example, among cells containing 500–1K UMI, 71% of SPG cells lack ChrY transcripts, whereas 53% of round spermatids lack ChrY transcripts (Figure S1F). Similarly, while only 10% of SPG cells lack ChrX transcripts, only 0.2% of round spermatids lack ChrX transcripts. This contrast is consistent across all UMI-bins examined, suggesting that round spermatids may indeed employ an active cytoplasmic sharing mechanism.

Third, early electron microscopy studies have shown that Sertoli cells undergo cyclic changes in the volume and surface area of their various organelles, including mitochondria, across the different stages of the seminiferous epithelial cycle (Ueno and Mori, 1990). Interestingly, when examining the percentage of mitochondria-encoded RNAs (%Mito) we find the highest and most variable levels of Mito RNAs are among the Sertoli and myoid cells (Figure 1D), suggesting that changes in mitochondrial morphology may correlate with mitochondrial transcriptional output. The levels of mitochondrial-encoded RNAs are much lower in the differentiating germ cells than in spermatogonia and somatic cells (Figure 1D), consistent with their reduced mitochondrial DNA copy number (Rantanen and Larsson, 2000).

Finally, earlier bulk RNA-seq analyses of germ cell populations have shown that, at a given sequencing depth, SCytes and STids have more detectable transcripts than somatic cells or other germ cell populations (Hammoud et al., 2014; Soumillon et al., 2013). What remains unclear is whether this higher number of genes detected in bulk SCyte and STid data was due to (1) a mixture of more heterogeneous cells or (2) each cell truly expressing a large number of distinct genes. Consistent with the latter, our single-cell data also show that SCytes and STids tend to contain a larger number of observed transcripts when compared to somatic cells or other germ cells (Figure 1D), leading to a larger “cell size factor” (a cell-

specific scaling factor that is proportional to the total number of unique transcripts observed per cell). To compare the transcript distribution properties across cells and major cell types, we then calculated a Gini Index for each cell, which is a measure of gene expression inequality *within* a cell. Here the property of even-ness, or equality, describes whether the highly and the lowly expressed transcripts cover a moderate range or a very wide range. For example, a cell expressing many distinct genes (high diversity) may express most of them at very low levels, and thus display high uneven-ness. Cells with such a severely uneven distribution of transcripts among genes will have a Gini closer to 1, whereas those with a more even distribution of transcripts will have a smaller Gini. For sparse count data, the classic Gini Index has inherent dependencies with the cell size factor. After correcting for this effect, we observed systematic differences across the 11 cell types. During the transition from spermatogonia to elongated spermatids there is a progressive increase in the Gini index (Figure S1G), suggesting that the germ cells in later stages show greater *inequality*, devoting a higher fraction of their transcripts to a narrower set of unique genes, which likely reflects the focus on increasingly specific biological functions (Piras et al., 2014; Teschendorff and Enver, 2017), even if there are more distinct transcripts observed in these cells. In sum, these analyses underscore the power of single-cell profile data, which can be used to compare the biological state within and between differentiating germ cells and somatic cells at all levels, from individual genes to whole-cell heuristics. As the community curates specific gene lists to represent additional functional processes, other biological properties, such as cell cycle, stem-ness, senescence, and migration, can also be scored for individual cells, further adding to the richness of information contained in this cell atlas.

Germ cell development includes initial discrete states followed by a continuous differentiation trajectory.

Constant sperm production relies on spermatogonial stem cells undergoing spermatogenesis asynchronously. Therefore, within a given cross section of a tubule (a snapshot of time in the seminiferous epithelial cycle) one finds multiple germ cell types spanning different stages of differentiation. This spatiotemporal complexity has made it challenging to isolate stage-specific cells with sufficient accuracy to decipher the developmental programs and their molecular drivers. Of the ~35K cells in our study, 20,646 correspond to germ cells with >1K detectable genes (range per cell 1–10K genes) (Figure 1B), allowing us to systematically identify distinct cellular states and key developmental transitions. The cell-cell distance matrix among the ~20K germ cells reveals cellular heterogeneity within and across clusters (Figure S2A). Unsupervised clustering of these cells identifies 12 germ cell states (GC1–12) (Figure 2A, Table S3A). The sequencing depth for each cell doesn't affect germ cell clustering (Figure S2B), but varies the cells placement along the path or within a cluster. For example, in each cluster we have a 10-fold range in cell size (number of genes detected per cell), with large and small cells coexisting in every segment along the continuous trajectory, which strongly suggest that minimum coverage we have in this dataset is sufficient to classify and position cells along the trajectory.

From these clusters, we find GC1–3 are discrete cell types, whereas GC4–12 follow a long, continuous trajectory, describing a smooth progression without distinct stable states separated by sparsely occupied transient states. The pattern of discrete and continuous

developmental transitions is also substantiated by the rank correlation matrix of the 12 germ cell states' centroids (Figure 2B). The developmental ordering of the 12 germ cell states was concordant with the pseudotime ordering from *Waterfall* (Shin et al., 2015) and *Monocle* (Trapnell et al., 2014) (Figure S2C, D).

To roughly define mitotically active, pre-meiotic, and post-meiotic cell populations within these 12 GC states, we calculated a mitotic cell cycle index for each cell, defined as the fraction of observed transcripts accounted for by 590 known cell cycle genes (Macosko et al., 2015). By this metric, cells in GC1 have the highest levels of mitotic cell cycle activity (Figure S2E), followed by gradually reduced levels in GCs 2–8, and vanishing levels in GCs 9–12. These observations are consistent with GC1 being spermatogonia, GCs 2–8 being meiotic cells and 9–12 being STids / ES. Our rough division between mitotic, meiotic and post-meiotic cell types is further corroborated by known marker genes (Table S3A). Cells in GC1 express genes such as *Zbtb16*, *Sall4*, *Gfra1*, *Sohlh1*, *Stra8*, and *Kit*, therefore, GC1 includes both undifferentiated and differentiating spermatogonia cells (note: PC1 vs. PC2 of all cells doesn't discriminate among SPG cells. See spermatogonia section for SPG-focused analyses) (Figure 2C, S2D). GCs 2–3 represent discrete developmental transitions and contain far fewer cells than other clusters. The relative rarity of these cells suggests that they correspond to transient cellular states *in vivo*, as it is unlikely that the underrepresentation of these populations can be readily attributed to selective loss during the experiments. These cells express genes enriched in RNA splicing and RNA binding proteins, early meiotic genes such as *Hormad1*, *Sycps*, γ -*H2Ax*, and *Dazl*, as well as chromatin remodeling and epigenetic modifiers, such as *Atr*, *Setx*, *Dnmt1*, *Chd*, *Brd*, *Ash1*, *Asxl2*, *Phf1/2*, *Mllt10* (*dot11*), and *Brd8/9*, suggesting that the transition from spermatogonia to early spermatocytes likely involves translational controls and changes in chromatin prior to entering meiotic prophase (Figure 2C, S2D). GCs 4–8 express mRNAs functioning at various stages of meiosis (Figure 2C, S2D). Although meiotic proteins have very defined stages of action histologically, we find that the RNAs of most of these genes are expressed broadly, over two or more clusters (Figure 2C), making it challenging to use known meiotic RNA markers to define meiotic stages precisely. To address this challenge we sought to develop new markers to define cell state along the germ cell developmental trajectory more precisely. Toward this goal we performed self-organizing map clustering, asking for an unbiased partition of cells into a linear series of 20 clusters. Strikingly, each of these clusters can be uniquely identified by 14–44 markers (mean 25.4), yielding a total of 508 markers that are transiently expressed in as narrow as one of the twenty clusters (Figure S2F). The list of ~500 markers genes may serve as relevant landmark genes that can precisely establish an unsupervised spatial map, and these novel markers can be used to distinguish meiotic and postmeiotic states more finely when combined with smFISH (Figure S2F, Table S3B). When validated, these genes may provide a resource on par with previous landmark datasets for different phases of mitosis (Whitfield et al., 2002).

Finally, GCs 9–12 express genes involved in acrosome formation or spermiogenesis (*Acrv1*, *Prms* and *Tnps*) (Figure 2C, S2D) and are also enriched for genes involved cytoskeleton organization and nuclear reorganization (Table S3A), consistent with these cells completing spermiogenesis to produce spermatozoa.

In short, our systematic analyses of the ~20k germ cells provides an unequivocal view of the continuous germ cell intrinsic program in the adult testis, and provides a list of stage specific markers that can finely partition the meiotic and postmeiotic process.

Dynamic gene expression regulation in germ cells identifies known and novel candidate regulators and suggests stage-of-action of infertility genes.

To delineate groups of genes showing concerted regulation over successive stages of germ cell development, we first calculated the average expression pattern (the “centroid”) of the 12 germ cell clusters, selected 8,535 highly variable genes across these 12 centroids, and used K-means clustering to partition these genes into six (Figure 3A) or 12 groups (Figure S3A). Interestingly, even with the unsupervised approach, these gene groups reveal a natural progression from genes highly expressed in germ cell clusters GC1 to those highly expressed in GC12 (Figure 3A). An alternative clustering method, self-organizing map (SOM), produced similar gene groupings, corroborating the dynamic regulation of distinct gene groups (Figure S3B).

The purpose of grouping genes by clustering is different from those seeking to identify a small number of highly significant stage-specific markers, as each gene group contains both the most specific markers and the rest of the genes showing a similar but less crisp pattern. For instance, in the six-group partition of the 8,535 genes, each group likely captures both the principle drivers of a given stage of development, but also the larger number of “follower” genes (Table S3C). This allows us to leverage these gene groups to gain further insights of the functional theme, regulatory network, and clinical consequence of germ cell development. First, gene ontology analysis of the six gene groups highlights the cascade of functional programs that are activated: Group 1 genes are enriched for those related to cell cycle, DNA repair, oxygen sensing and response, and oxidative phosphorylation (FDR <5%). These results are consistent with an actively dividing population of spermatogonia. Genes in Group 2 are enriched for RNA processing, RNA splicing, alternative splicing, and TGF- β signaling. TGF- β signaling was previously shown to be induced or augmented in response to hypoxia in multiple cell types (Zhang et al., 2003), and it initiates an epithelial mesenchymal transition - allowing cells the ability to acquire a migratory potential (Xu et al., 2009), suggesting that GC2 corresponds to cells in preparation for, or in the process of, crossing the blood-testis barrier. Finally, genes in Groups 3–6 represent downstream processes such as spermiogenesis and flagella formation.

To delineate major transcriptional regulators acting in individual stages we applied comprehensive motif discovery analyses using the putative promoter sequences (1 kb flanking each side of the transcription start site) of genes within each group (Bailey et al., 2009; Machanick and Bailey, 2011). Motif enrichment patterns in promoter sequences suggest that genes within each group are likely regulated by distinct sets of transcription factors (TFs), i.e. very few TFs are present in multiple gene groups. Many of the transcription factors enriched only in group 1 genes have either an established role in SPG development (i.e. *Bcl6b*) (Oatley et al., 2006) or are implicated in infertility in gene-knockout experiments (*Zfx*, *Nrf1*, *E2f* family, *Ctcf1*, and *Egr1*) (Figure 3B, Table S4A,B) (El-Darwish et al., 2006; Luoh et al., 1997; Suzuki et al., 2010; Wang et al., 2017). In

Author Manuscript

addition to TF categories described above, we find several motifs that either correspond to a known transcription factor that has not been previously explored in the testis (*Zbtb33*, *Runx3*, and *Zx1*), or very strongly enriched motifs that have no annotated transcription factor - providing us with lists of genes/motifs that need to be explored in future studies *in vivo*.

Author Manuscript

Finally, to explore the clinical utility of our single-cell data and the resulting gene-dynamics map of germ cell progression, we focused on ~200 previously reported mouse infertility genes (Matzuk and Lamb, 2008) and observed their dynamic expression patterns across the 12 cluster centroids. A large fraction of mouse infertility genes have peak expression in the spermatogonia and round/elongating spermatid stage; and this pattern holds true also for human infertility genes documented by OMIM (Figure 3C, Table S3D). We then asked if the timing of a gene's peak expression correlates with its observed stage of germ cell arrest once mutated. Interestingly, many genes causing arrest in one stage tend to be expressed at the highest levels in the same or an earlier stage. For instance, infertility genes that manifest as arrest at the spermatocyte stage tend to be expressed at highest levels in the preceding stage of spermatogonia. The observations of such a "lag" suggest that aberrant meiotic progression may manifest more according to the time of translational deficits than the time of transcriptional dysregulation. In contrast to meiotic arrests, the round and elongated spermatid arrests exhibit the expected concordance between expression timing and stages of arrest (Figure 3D). In short, we anticipate that these gene groups of distinct dynamic patterns can be used as a searchable resource to predict the stage-specific consequences of perturbing individual genes.

Single-cell data identify four spermatogonial subtypes that correspond to spermatogonial states previously described by histology.

Author Manuscript

Through a series of genetic experiments, many groups have independently identified a handful of spermatogonial stem cell markers that capture the developmental progression from undifferentiated (e.g., *Gfra1*, *Lin28*, *Id4*, *Pax7*, *Etv5*, *Zbtb16*, *Tert* and *Bmi1*) (Hara et al., 2014; Nakagawa et al., 2010); Chakraborty et al., 2014; (Chan et al., 2014; Sun et al., 2015); (Aloisio et al., 2014) (Chen et al., 2005); (Buaas et al., 2004; Costoya et al., 2004); (Komai et al., 2014) (Pech et al., 2015) to differentiating spermatogonia committing entry to meiosis (e.g., *Kit*, *Stra8*, *Dmrts*, and *Sohlh3*) (Anderson et al., 2008; Kissel et al., 2000; Matson et al., 2010; Suzuki et al., 2012; Xu et al., 2009; Zhang et al., 2014). However, little is known about the transcriptome-wide dynamics during the differentiation process, the finer steps of the process, or the transcriptional regulators that are important for each step.

Author Manuscript

Our dataset contains 2,484 spermatogonia cells, providing an excellent opportunity to re-examine these important questions. We performed focused re-clustering of 2,484 spermatogonia cells and identified 4 subtypes (Figure 4A, B). Each of the four subtypes is comprised of cells with a range of 1,000 UMI – 10,000 UMI, suggesting that the number of genes per cell has a minimal effect on spermatogonia subtypes (Figure S4A). Global transcriptome patterns and developmental ordering of cells suggest that the four SPG subtypes follow the order of SPG1 to SPG4. Marker gene analysis suggests that cells in SPG1 correspond to undifferentiated spermatogonia, as they express one or more

spermatogonial stem cell genes such as *Id4*, *Bcl6*, *Taf4b*, *Gfra1*, *Lhx1*, *Etv5*, *Eomes*, and *Plzf* (a.k.a. *Zbtb16*), and lack the expression of differentiation markers (Figure 4B, C, Table S5A). Therefore, we predict SPG1 (~213) cells is comprised of a mixture of A_{single} (a single spermatogonia stem cell), A_{paired} (two connected spermatogonia), and A_{aligned} (chains of 4, 8, 16, or occasionally 32 spermatogonia). Unlike SPG1 subtype, cells in SPG2–4 subtypes express various differentiation marker gene combinations (Figure 4B, C), suggesting that these three SPG subtypes represent progressively differentiated spermatogonia (Matson et al., 2010; Schrans-Stassen et al., 1999; Suzuki et al., 2012; Zhang et al., 2014; Zhou et al., 2008). Cells in SPG2 are *Kit*⁺ and *Stra8*⁺, express early differentiating markers such as *Dmrt1*, *Dmrtb1*, *Sohlh1*, *Sohlh2*, and lack any evidence for meiotic gene expression (Figure 4C). These patterns are consistent with SPG2 corresponding to A_{1-4} differentiating spermatogonia (Zhang et al., 2014; Zhang and Zarkower, 2017). Cells in SPG3 are *Kit*⁺, *Stra8*⁻, *Sohlh2*⁻, *Sohlh1*⁺, *Dmrtb1*⁺, *Dmrt1*⁺, and express meiotic genes (such as *Sycp3* or *Prdm9*) at very low levels. Specifically, the loss of *Sohlh2* and low levels of meiotic genes suggests that this population is consistent with $A_{\text{intermediate}}$ (A_{in}) – Type B spermatogonia (Suzuki et al., 2012). Cells in SPG4 are *Kit*⁺, *Stra8*⁺, and express high levels of meiotic genes, suggesting that SPG4 cells are consistent with Type B or preleptotene cells which are poised for meiotic entry. Notably, the off-on-off-on pattern of *Stra8* mRNA expression across the four populations initially seemed unexpected (Figure 4C), but this bi-phasic activation of *Stra8* across SPG states becomes clear when considering spatial positions of SPG2 (A_{1-4}) and SPG4 (type B/Prelep) cells in the seminiferous tubule (Figure 4D). In stages VII–VIII of the seminiferous epithelial cycle, SPG2 and SPG4 populations are coincident, therefore, both are exposed to retinoic acid (RA) during this defined developmental window. The responsiveness of cells in SpG2 and SpG4 to RA is consistent with earlier data that has shown that retinoic acid signaling or dietary supplementation of RA precursors (vitamin A) is necessary for the transition of undifferentiated spermatogonia (A_{undiff} (SPG1) to A_1 (SPG2) spermatogonia), and the transition of preleptotene cells (SPG4) into meiosis (Anderson et al., 2008; Endo et al., 2015; Hogarth et al., 2013; Morales and Griswold, 1987; Snyder et al., 2011; Van Beek and Meistrich, 1990). Interestingly, although both developmental transitions require active RA signaling, RA induces distinct cell type specific gene expression and exerts stage-specific outcomes, underscoring the importance of transcription factor repertoire or chromatin context of the cells in determining the signaling outcome.

Finally, to better define the complex networks of transcription factors that may be involved in coordinating the developmental progression of spermatogonia, we identified 57 subtype-specific TFs over the four SPG states (Table S4C), 12 of which are shown in Figure S4B. Some of these TFs are specifically expressed in one state (e.g., *Egr1* in SPG1; *Lmo1* and *Tcea3* in SPG2; *Tead2*, *Esx1*, and *Pthf1* in SPG3; and *Nr2c2*, *Nfat5*, and *Hif1a* in SPG4), whereas others span multiple states (e.g., *Esx1* in SPG3–4 and *Cited* in SPG2–4), suggesting broader functional activity. Taken together, this analysis provides the first molecular signatures of the major spermatogonia subtypes described histologically and identifies candidate transcriptional regulators that may act in a single state or across multiple states.

The undifferentiated SpG1 spermatogonia population lacks distinct molecular states.

Early pairwise immunohistochemistry staining of many spermatogonial stem cell markers *in vivo* found that most identified stem cell genes show uneven distribution of markers among and across the undifferentiated spermatogonia populations comprised of A_{single} , A_{paired} , and A_{aligned} spermatogonia (de Rooij, 1998; de Rooij and Russell, 2000; Nishimune et al., 1978; Ohbo et al., 2003; Shinohara et al., 2000). In order to achieve a more global sense of the heterogeneity within the undifferentiated spermatogonia cells, we performed a focused analysis of the 213 cells in SPG1. PCA and Louvain-Jaccard clustering using all genes, highly variable genes, or only the established stem cell genes (that can distinguish the A_{single} , A_{paired} , and A_{aligned} cells among the undifferentiated SPG cells) each consistently identified a continuous ensemble of cells without discernible stable structure (Figure S4C). When forcing the data into two clusters, the three gene sets brought two-cluster solutions that were completely discordant (Figure S4C). This result suggests that SPG1 cells in our dataset do not reveal distinct functional subtypes and thus do not support the hierarchal model within undifferentiated SpG. Rather, our single cell sequencing findings are consistent with a model of spermatogonial stem cell plasticity previously described (Hara et al., 2014). However, it is also possible that the developmental hierarchy among the undifferentiated SPG cells is maintained at the level of protein content or cell-cell interaction, or alternatively, is dependent on very subtle transcriptomic differences which would require much deeper sequencing to discern. Nevertheless, our results yielded specific gene expression markers for SPG1 and can be leveraged as new reagents for SPG1 enrichment for future in-depth analysis, or for *in situ* tracking to unravel the lineage relationships and spatial complexity.

Identification of known and novel somatic cell populations.

Successful spermatogenesis in mammals requires the support of a specialized microenvironment (i.e., the niche) consisting of diverse somatic cell types. In the past, the major somatic cell types were identified histologically, and their functional roles have been determined using genetic strategies (DeFalco et al., 2015; Yoshida et al., 2007). However, a comprehensive census of major cell types in the somatic compartment has been hampered by their relative rarity: unselected cell isolation from the testis tend to recover too few somatic cells. Here, we applied molecular and genetic strategies to enrich for single cells from the somatic compartment. Among the cells analyzed in these targeted enrichment experiments, ~5,000 can be assigned to somatic cells. Clustering analysis focusing on these cells revealed seven major cell types (Figure 5A). Five of the clusters were recognized as known cell types based on previously reported cell type-specific marker genes: Sertoli cells (*Sox9*), Leydig cells (*Hsd3b1*), myoid cells (*Acta2*), endothelial cells (*Vwf*), and macrophages (*F4/80* a.k.a *Adgre1*) (Figure 5B). The two remaining clusters represent unexpected populations, corresponding to (1) an innate lymphoid type II immune cell type not known to be present in the testis and (2) a previously unknown mesenchymal cell population (further described below).

Cells in the first unexpected cluster have high expression levels for *Id2*, *Gata3*, *Cd90*, *IL7R*, *IL13*, and *Rora* (Table S2), which are cell surface, intracellular, and cytokines markers characteristic of innate lymphoid type II (ILCII) cells – a population similar to T-helper cells

(T_H cells)(Spits and Di Santo, 2011). These cells were first described in the spleen, mesenteric lymph node, and bone marrow of mice, where they play a role in regulating immune responses (Neill et al., 2010). However, these cells were not previously looked for, nor known to exist, in the testis. To confirm the presence of this population in the testis by an independent approach, we performed whole-animal perfusion with 1X PBS in order to clear circulating immune cells from the testis, and then dissociated and stained the single suspensions using a variety of surface markers (CD45, THY1, CD3, CD4, CD8, IL7R), intracellular transcription factors (GATA3) or immune cytokines (interleukin 4 (IL-4) or IL-13) (see methods, Figure 5C, D, S5A). Immune cell profiles from total testis show that the total percentage of immune cells (CD45⁺) in the testis is ~8% (Figure S5A). Of the CD45⁺ cells, ~3% are Thy1.2⁺ (Figure S5A), which marks both T-helper cells (T_H2) and ILCII populations. To distinguish between these populations, we first gated cells based on CD3 expression (Figure S5A) followed by CD4, CD8, and IL7R (Figure 5C, S5A). Based on these flow profiles we can conclude that the CD3⁺, CD4⁺, CD8⁻, and IL7R⁺ cells are T_H2 cells, whereas the CD3⁻, CD4⁻, CD8⁻ and IL7R⁺ cells are ILCII. To further verify that the triple negative population (CD3⁻, CD4⁻, CD8⁻, IL7R⁺) is truly an ILCII population, we stained for a panel of intracellular transcription factors and cytokines and found that the testis ILCII population is positive for Gata3 and IL-13, but not IL-4 (Figure 5D).

Cells in the second unexpected cluster show high expression levels for *Tcf21*, *Arx*, *Vim*, *Col1A1*, and *Sca1*, and are recognized as a mesenchymal cell population (Figure 5B, Table S2). To determine the location of these cells in the testis, we used a genetic labeling strategy (*Tcf21-creERT2; tdTomato*), which confirms the presence of a Tcf21⁺ cell population surrounding the seminiferous tubules (Figure 5E) and interstitial space. Furthermore, we molecularly enriched for the Tcf21⁺ cell population by flow sorting cells with the surface markers Sca1⁺/Kit⁻ from the testis (Figure 5F). This population has high levels of *Tcf21* and *Col1A1*, but lacks expression of *Myh11*, *ActA2*, indicating that this population is distinct from both mature myoid and Leydig cells.

To determine whether a similar population is detected in the embryonic gonad population, we reprocessed the previously published single-cell RNA-seq data for *Nr5a1*-GFP⁺ cells in XY mouse gonads during sex determination (Stevant et al., 2018), which identified six somatic cell types. Comparisons with the seven somatic cell types found in our study demonstrate that the adult endothelial, Leydig, and Sertoli cells show high correspondence to cells in the embryonic gonad tissue, while our unknown cell type is most similar to the interstitial progenitor cells in the embryonic gonad (Figure S5B). Interestingly, pseudotime ordering of this embryonic population by Stevant *et. al.* suggests that the interstitial progenitor population may give rise at least to fetal Leydig cells. Whether the adult Tcf21⁺ population acts as a reserve somatic progenitor in the adult during tissue homeostasis or tissue regeneration remains to be determined.

Next, we compared our data with those reported for the Mouse Cell Atlas (MCA) which analyzed >16K cells from the adult mouse testis (Han et al., 2018). While our atlas contains enhanced representation of somatic cell types, the MCA is dominated by germ cells, and lacks dense survey of other cell types. As a result, unsupervised clustering could not reliably identify major cell types within the MCA. By a semi-supervised approach using most

definitive markers developed from our study, we were able to reach a provisional identification of seven minor cell types in MCA aside from the germ cells, which are the overwhelming majority. After calculating the rank correlation between each of the 8 cluster centroids from MCA with each of the 11 major cell type centroids in our study, we found that the Leydig cells, myoid cells, macrophages, and Sertoli cells can be identified in MCA, whereas our unknown, endothelial cells, and the innate lymphoid cells were not found (Figure S5C). Furthermore, there is no discernible substructure among the MCA germ cells that correspond to spermatocytes, the round spermatids, or elongated spermatids.

Taken together, the analysis of a large number of single somatic cells identifies known and rare, unexpected populations in the testis. Future studies will be needed to characterize the functional significance of the ILCII and mesenchymal cell populations in testis development and tissue homeostasis.

Sertoli subtypes capture transcriptional changes across the stages of the seminiferous epithelial cycle.

Sertoli cells are the only somatic cells within seminiferous tubules that intimately interact with developing germ cells (Figure 6A). As a result, the functional role of Sertoli cells has been a subject of intense investigation for decades. Sertoli cells residing at different stages of the seminiferous epithelial cycle exhibit characteristic differences in size, shape and marker gene expression patterns (Hasegawa and Saga, 2012; Johnston et al., 2008; Kerr, 1988a, b). These features (Figure 6A) have supported the description of Sertoli cell heterogeneity in terms of 12 spatially-ordered stages of the seminiferous tubule (labeled I-XII in Figure 4D).

Our atlas allows the first direct analysis of a large number of Sertoli cells from the adult testis (Figure 5A), isolated from two transgenic lines *Sox9-eGFP* and *Amh-cre;mTmG*. Unsupervised clustering of ~1,100 Sertoli cells with >1,000 detected genes identifies four stable cell clusters (SER-1 to SER-4) which can be further divided into nine sub-clusters, denoted SER-1, SER-2A/B, SER-3A/B, and SER-4A/B/C/D, where the A-B split indicates finer divisions among transcriptomically similar cells (Figure 6B). The nine subtypes showed different functional attributes (Figure S6A) and all nine were observed in both of the two transgenic lines (Figure S6B, C). A natural question then is how the four types of Sertoli cells, or the nine *subtypes*, match to the histologically defined *stages*. We took advantage of previously reported stage-specific marker genes identified by microarrays from tubule segments (Hasegawa and Saga, 2012; Wright et al., 2003) and compiled them into four lists, corresponding to the mouse seminiferous tubule stages I-III, IV-VI, VII-VIII, and IX-XII, respectively. We then calculated the relative “loading”, or correspondence (see Methods), of the four curated lists of stage-specific genes in each of the nine sub-clusters observed in our data (Figure 6C). Interestingly, we do not find a simple one-to-one matching between the four histological stages and the four major clusters. Rather, cells within the same major cluster, such as SER-2A/B, may correspond equally well to two different stages; or, equivalently, cells with the same stage-specific signatures may appear in multiple clusters, such in both SER-2A and 3A. More specifically, cells in SER-2A express genes associated with Stages IX-XII, while cells in SER-2B, although computationally predicted to be closely related to SER-2A, map to Stages IV-VI (Figure 6C). Therefore, the single-cell analysis of

Sertoli cells uncovers a fundamental difference between transcriptome-wide functional properties of Sertoli cells and spatially defined developmental stages of the seminiferous tubule. This result implies that 1) clustering of cells are influenced by all active biological pathways and global cell attributes, but these global attributes may not fully reveal specific programs that confer positionally-defined biological stages; and 2) functionally dissimilar Sertoli cells – as defined by global transcriptome patterns – may co-localize *in situ* and serve complementary functions.

To independently validate the predicted spatial positions of the 9 Sertoli cell subtypes, we performed single-molecule fluorescent *in situ* hybridization (smFISH) or single molecule hybridization chain reaction (smHCR), followed by lectin immunostaining. The shape of the lectin distribution in individual germ cells (shown in green) shows characteristic patterns that allow us to estimate the stage of the seminiferous epithelial cycle for each tubule cross section (Figure 6E, S6D, S7A,B). However, a challenge to image RNA transcripts *in situ* in intact Sertoli cells are at least two-fold. First, Sertoli cells are large - they traverse the tubule radius and are in direct contact with, and often enclose germ cells. Second, we lack cell surface markers that define the perimeter of the Sertoli cells. To overcome these challenges, we focused our smFISH and smHCR analysis on a series of highly variable Sertoli cell subtype-specific marker genes that are absent in germ cell populations (Table S6). Therefore using this approach, we feel it is reasonable to deduce that the observed puncta from either smFISH or smHCR are part of the Sertoli cells of a specific stage (that lectin-staining has defined), and not part of the germ cells.

The marker gene lists identified for the 9 Sertoli cell clusters contain many of the previously described stage-variable genes (i.e. *Gas6*, *Drd4*, *P2rx2*, *Zfp3611*) identified from tubule segments isolated by transillumination (Hasegawa and Saga, 2012; Wright et al., 2003), as well as newly discovered subtype marker genes such as *Mfge8*, *Prm2*, *Lgals*, *Caskin1*, *Ptprv*, *Mical2*, *Eyst3*, *Zfp3611*, and *Dpysl4*. Importantly, the majority of marker genes selected for smFISH or smHCR recapitulated the predicted Sertoli subtype stage specific expression patterns (Figure 6C,D, Table S6). For example, previous literature and our Drop-seq data both predict that *Gas6* is highest in I-III, dropping in IV-VI and to the lowest point in VII-VIII, and partially recovers in IX-XII (Figure 6C, Table S6), which is indeed observed in our smFISH patterns (Figure S6D) and quantification (Figure S7D). Similarly, the predicted patterns were seen for genes such as *Drd4*, *Dpysl4*, *Mfge8*, *Pr2x2*, *Eyst3*, *Zfp361*, and *Zfyve27* (Figure S7C, D). Further, to expand the number of genes tested simultaneously in a single tissue cross-section, and have a better sense of presence/absence of signals across stages, we performed multiplex sequential smHCR using a panel of markers that can distinguish between the Sertoli cell subtypes. Specifically, by using a combination of markers (*Lagls*, *Caskin1*, *Ptprv*, *Mical2*, *Eyst3*), we confirm that cells derived from a major Sertoli cell cluster (Ser-4C/D vs. Ser-4A/4B, or Ser-3A vs. 3B) reside in distinct locations of the seminiferous epithelium (Figure 6D-E).

A particularly unexpected marker of Sertoli cells was protamine 2 (*Prm2*) - a sperm specific nuclear protein highly expressed in the round spermatid, but under translational control (Table S6). smFISH validation of *Prm2* transcript in the testis (Figure S7A,B) shows protamine signal in regions surrounding the round spermatid nuclei (stages VII-VIII) and in

Sertoli cell tips. This cytoplasmic staining pattern of *Prm2* transcript is observed in Sertoli cells of almost all stages of the seminiferous tubule cycle (except VII-VIII). The persistence of protamine RNA is surprising because one expects that RNAs brought in from phagocytosed germ cells or residual bodies should be degraded immediately by the lysosome. Instead, these RNAs persist in Sertoli cell cytoplasm (Figure S7A, B). To determine and validate the site of transcription, we designed a protamine probe set containing intronic and UTR sequence, therefore allowing us to capture transcriptional foci in the nucleus. As expected, *Prm2* transcriptional foci are detected in the round spermatid nucleus at stages VII-VIII, whereas, the nuclear signal is absent from Sertoli cell nuclei, despite having *Prm2* transcript in the cytoplasm - which is consistent with the notion that the *Prm2* RNAs present in Sertoli cell cytoplasm are persistent RNAs of germ cell origin. If and whether a biological role exists for these retained RNAs remains to be examined.

In short, by using both known and novel markers for each Sertoli subtype, we have successfully linked the 9 Sertoli cell subtypes to the predicted stages of the seminiferous epithelial cycle (Figure 6C-E, S6D, S7C-D, see Methods), and showed that many of the selected Sertoli subtype markers are not regulated in an on-off manner, underscoring the continuous nature of Sertoli cell progression.

Discussion

Spermatogenesis is characterized by three specific functional phases: mitotic proliferation and expansion, meiosis, and spermiogenesis. In the proliferation phase, spermatogonia (SPG) lining the basement membrane asynchronously undergo several mitotic divisions to form spermatocytes (SCytes), which then complete two meiotic divisions to form haploid spermatids (STids). The STids proceed through the process of spermiogenesis, which entails morphological, structural, and chromatin changes, ultimately giving rise to mature sperm. This differentiation process is coordinated radially within a tubule cross-section, and occurs asynchronously along the tubule. As a result, at any point along the seminiferous tubule there are multiple differentiating germ cells at different stages of development. The longitudinally continuous and radially asynchronous process of spermatogenesis has made it challenging to obtain stage specific molecular resolution of germ cell development. As a result, past studies relied on histological descriptions (Clermont, 1972), characterized transcriptomes of purified cells using known cell surface markers (Guo et al., 2017; Hammoud et al., 2014; Hammoud et al., 2015; Lesch et al., 2013; Oatley et al., 2006; Yoshida et al., 2007), or analyzed the semi-synchronous or artificially synchronized first wave of spermatogenesis in the neonatal testis (Ball et al., 2016; Zimmermann et al., 2015). Therefore, the field of stem cell, regenerative, and reproductive biology still lacks a comprehensive catalogue of major cell types, cell states, associated molecular markers and signaling pathways that guide this developmental process. We addressed this challenge by performing single-cell RNA-seq analysis of ~35,000 cells of the adult testis.

A complete cellular and molecular catalog of spermatogenesis.

At a gross level, we identified all major germ cell groups including spermatogonia (stem and progenitor populations), spermatocytes, round spermatids, and elongating spermatids.

Unsupervised ordering of cells allowed us to reconstruct a complete differentiation trajectory of the spermatogenic process at an unprecedented resolution (Figure 2, 7; GC1-GC12). Our annotation of germ cell subtypes (GC1-GC12) uses an existing set of genes from prior knowledge. However, many of these genes exhibit broad expression patterns and may not be the most specific molecular markers for individual stages. Instead, we generated a list of tightly regulated genes along the germ cell differentiation trajectory to precisely define subtypes, but these markers need validated by smFISH in order to provide a highly detailed molecular map for meiosis – a resource on par with previous landmark datasets for different phases of mitosis (Whitfield et al., 2002).

Our data demonstrate for the first time the continuous nature of germ cell development with no stable intermediate states. Although spermatogenesis appears largely continuous, there is a rare and discrete developmental transition (captured by clusters 2/3) occurring prior to entry into meiosis. The cells in these clusters express multiple transcriptional cofactors, epigenetic modifiers, and remodelers including *Setx*, *Dnmt3a*, *Cbx1*, *Kdm5a*, *Ash1*, *Asx12*, *Phf1/2*, *Mllt10 (dot1)*, and *Brd7/8/9*. Previous genetic labeling or loss-of-function experiments demonstrate a role for some of these factors in germ cell development. For example: 1) *Setx*^{-/-} mice exhibit a severe disruption of the seminiferous tubules and early meiotic arrest in 35-day males (Becherel et al., 2013); 2) Analysis of postnatal testis in *Asx12* gene trap mice shows that *Asx12* expression is restricted to early spermatocytes and is not detectable in secondary spermatocytes. Full body knockout of *Asx12* results in early neonatal lethality, therefore, fertility could not be evaluated (Baskind et al., 2009); 3) PHF1 protein, which is comprised of an N-terminal Tudor domain and two C-terminal PHD fingers, play important roles in Polycomb repressive complex 2 (PRC2)-mediated transcriptional repression through stimulating H3K27me3 activity by binding to H3K36me3 (Cai et al., 2013; Musselman et al., 2012; Qin et al., 2013). A recent investigation showed that PHF1 binds to H3K27me3 on a testis-specific H3 variant (Kycia et al., 2014), suggesting that some well-studied somatic epigenetic “readers” might play distinct but yet-to-be-identified roles specifically in germ cells.

While the initial analysis of all cells identified spermatogonia cells, we could not clearly distinguish SPG subtypes. Only by focused re-clustering of the ~1,200 SPG cells could we resolve the undifferentiated and differentiated spermatogonia. This analysis discerns four SPG subtypes that correspond to spermatogonial stem cell populations previously described histologically (Figure 4,7). A further zoomed-in analysis of the undifferentiated SPG1 population doesn't reveal any structural hierarchy or stable states, suggesting cellular plasticity at the RNA level consistent with stem cell plasticity model of the undifferentiated SpG cells.

In addition to reconstructing comprehensive developmental maps and gene expression networks, we have identified both known regulators (*Egr1/4*, *Bcl6b*, *Nrf1*, *E2f4*, *Nfyb*, *Ctcf1*, *Rfx2*) (Danielian et al., 2016; Fukuda et al., 2013; Sleutels et al., 2012; Tourtellotte et al., 2000; Wang et al., 2017) and previously undescribed regulators (*Zbtb33*, *Zbtb7a*, *Rfx3/4*, *Runx3*) in germ cell development (Figure 3B), which will need to be further validated using molecular and genetic approaches. Additionally, we find a large number of uncharacterized motifs in both meiotic and postmeiotic cells – which raises the possibility of identifying

gametogenesis-specific factors that were missed from somatic cell ENCODE datasets. In sum, the cellular catalog, developmental trajectories, transcriptional programs and candidate regulators described here deepened our understanding of the gametogenesis process and far exceed the level of granularity described in earlier foundational datasets performed on sorted bulk populations (Guo et al., 2017; Hammoud et al., 2014; Lesch et al., 2013; Soumillon et al., 2013).

Characterization of the somatic compartment of the testis

Although the enrichment strategies, by definition, led to a cellular census that no longer reflects the true proportion of cell types *in vivo*, our approach gained the advantage of efficiently charting the entire molecular landscape of testis somatic cells. The markers we developed for both common and rare cells will enable future rounds of spatiotemporal mapping of cell-cell interactions/communications in a complex organ structure. Even at this early stage of map building, we made the provocative discovery of two unexpected somatic cell types.

The first is an ILCII population, which is related to T_H2 cells in terms of cytokine production, and have been studied in the context of intestinal homeostasis (Yang et al., 2016). The ILCII population in the testis is capable of secreting high levels of IL-13, but not IL-4 *in vitro* (Figure 5). The presence of IL-13 in the testis has been detected using real-time RT-PCR, but the source was not known (Maresz et al., 2008). However, the authors of that study nicely demonstrate that the IL-13 cytokine and receptor are necessary for maintaining an alternative subpopulation of macrophages known as M2 ($ym1^+$) macrophages (Maresz et al., 2008). Based on these preliminary observations from the testis and from other tissues, we postulate that the ILCII population in the testis may have an immune surveillance and tissue homeostasis function.

The second population identified was a $Tcf21^+/Sca1^+$ population. Based on the transcriptome data, comparative analysis with earlier single cell transcriptomes, and gene ontology analysis we predict that the $Tcf21^+$ cells is a mesenchymal cell population that is reminiscent of an embryonic interstitial cell progenitor population has been previously predicted to give rise to fetal supporting cells by pseudotime ordering (Acharya et al., 2011; Cui et al., 2004; Stevant et al., 2018). However, the role of this population in the adult testis is unknown, and whether this population can serve as reserve somatic stem cell remains to be tested. Future studies will aim to elucidate whether the ILCII and/or $Tcf21^+$ populations play an essential role in testis tissue homeostasis.

Finally, for somatic cell populations with sufficient cell numbers, such as the Sertoli cells, we applied iterative clustering to uncover previously unappreciated finer-level heterogeneity. Re-clustering of Sertoli cells identified four major subtypes that can be further divided to nine molecular clusters (Figure 6, 7). We found that each major cluster can contain multiple functional types; whereas conversely, each functional type of Sertoli cells could reside in multiple stages of the seminiferous epithelial cycle. Furthermore, the cell type-specific markers we produced, in conjunction with stage-specific markers, will be invaluable for dissecting the co-localization patterns of diverse classes of Sertoli cells, unraveling the

functional heterogeneity of Sertoli cells within a single cross-section, and resolving germ cell – Sertoli cell communication.

Taken together, our datasets and findings will likely serve as an enduring resource to the community, and are critical for an integrative understanding of germ cell development and germ cell – niche communication. Such an understanding represents an essential step toward finding new ways to recapitulate this process *in vitro* in the context of developing novel therapeutics for infertility.

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by the corresponding author S. Sue Hammoud (hammou@med.umich.edu). Questions regarding computational resources can be directed to co-corresponding author Jun Li (junzli@med.umich.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All animal experiments were carried out with prior approval of the University of Michigan Institutional Committee on Use and Care of Animals (Animal Protocols: PRO04380, PRO06792), in accordance with the guidelines established by the National Research Council *Guide for the Care and Use of Laboratory Animals*. Adult (7 to 18 weeks old) male mice were housed in the University of Michigan animal facility, in an environment controlled for light (12 hours on/off) and temperature (21 to 23°C) with *ad libitum* access to water and food. For detailed mouse strain information, see below and Key Resources table.

METHOD DETAILS

RNA-sequencing

Isolation of mouse cells for sequencing: Testes from adult C57BL/6 (JAX@mice, stock no. 000664) and transgenic mice were excised and the tunica albuginea was removed. Briefly, seminiferous tubules were transferred to 10ml of digestion buffer1 (comprised of Advanced DMEM:F12 media (Life Technologies), 200 µg/ml Collagenase IA (Sigma), and 400 units/ml DNaseI (Worthington Biochemical Corp.)). Tubules were dispersed by gently shaking by hand, and allowed to settle for 1 min at room temperature. Tubules were then transferred to digestion buffer 2 (200 µg/ml trypsin (Invitrogen) and 400 units/ml DNaseI (Worthington Biochemical Crop) dissolved in Advanced DMEM:F12 media) and dissociated at 35°C / 215 rpm for 5 min each and quenched with the addition of fetal bovine serum (FBS). Cells were filtered through a 100 µm strainer, washed in Phosphate-buffered saline (PBS), pelleted at 600g for 3min, and re-suspended in MACS buffer containing 0.5% BSA (MACS buffer; Miltenyi Biotec). For interstitial cell enrichment, testes were dissociated in digestion buffer 1 (described above) for 5min at 35°C / 150 rpm, the cells were dislodged by a gentle hand shake, supernatant quenched with FBS (tubules were discarded), and used directly for Drop-Seq. For all Drop-seq experiments, live single-cell suspensions were collected by flow cytometry using FACSARIA II/III (BD Biosciences) and Synergy SY3200 (Sony) cell sorters.

Enrichment of rare cell populations: Because cells such as spermatogonia, interstitial and Sertoli cells are very rare and proportionally less well-represented in the unbiased six seminiferous tubule (ST) dataset (6 batches), we employed a number of genetic or molecular strategies to enrich for these various cell types. Specifically, we did 1n-depletion experiment to remove haploid round and elongating spermatids (2 batches), which account for >50% of the cells of the testis and this naturally increases the representation of rarer cells. We also specifically enriched for undifferentiated and differentiating spermatogonia (SPG) (3 datasets), Interstitial (INT) cells (6 datasets), Sertoli (SER) (8 datasets) in order to achieve a more comprehensive census of major cell types testis (Table S1A).

In-Depletion: Hoechst 33342 (Life Technologies) and propidium iodide (PI) staining was performed on single cell suspensions of testis as previously described (Gaysinskaya et al., 2014). This method allows us to identify and remove haploid germ cell subtypes, while maintaining all other testicular cells.

SPG-enrichment: Spermatogonia were enriched by selection of C-kit+ or Gfra1+ cells. C-kit+ (CD117) cells were isolated from whole tubule cell suspensions on a magnetic cell-sorting separator (Miltenyi Biotec) using an anti-CD117 (Miltenyi Biotec) antibody. Cells were additionally stained with a biotinylated anti-CD117 antibody (1:200) for 20 min followed by streptavidin conjugated Alexa Fluor 488 (1:1000, Life Technologies) for 20 min prior to flow cytometry. For Gfra1+ selection, tamoxifen inducible *Gfra1^{CreERT2}* mice on a C57BL/6 background (kindly provided by Dr. Shosei Yoshida, National Institute for Basic Biology, Okazaki, Japan) were crossed with B6. *Gt(ROSA)26Sor^{tm4}(ACTB-tdTomato,-EGFP)Luo* (*Rosa^{mT/mG}*; JAX®mice, stock no. 007676). Labeling of Gfra1+ spermatogonia was conducted by injecting a *Gfra1^{CreERT2}; Rosa^{mT/mG}* mouse with 2mg of 4OH-tamoxifen (dissolved in ethanol and then in corn oil, Sigma) per day for 3 weeks prior to euthanasia at 8 weeks of age. All live Gfra1 positive (GFP expressing) and tdTomato negative cells were collected by flow cytometry. The extended labeling was necessary to obtain the number of cells needed for Drop-seq. As a result this dataset includes spermatogonia and spermatocytes.

Interstitial enrichment: Interstitial cell dissociations were performed by modifying digestion buffer 1 – which replaces 2.5mg of collagenase 1a1 with 0.25mg of collagenase D (Sigma). This milder digest is sufficient to gently dissociate the interstitial cells, but the second digestion buffer containing trypsin is the same for all sample preps since this is required to dissociate the seminiferous tubule into a single cell suspension. For Thy1+ enrichment, we performed a magnetic cell-sorting separator (Miltenyi Biotec) using anti-CD90.2 (Miltenyi Biotec) antibody. Cells were additionally stained with anti-CD90-FITC conjugated (1:200; Abcam, Cat# ab62009; RRID: AB_940927) or anti-CD90-PerCP cyanine5.5 (1:200; Life Technologies, Cat# 45090082; RRID: AB_2573662) for 20 min prior to flow cytometry and CD90+ cells were collected for Drop-seq. For Sca1+ (Ly6a) cell enrichment, cells were stained with anti-Ly6a-PerCP cyanine5.5 (1:200; Life Technologies, Cat# 45-5981-82; RRID: AB_914372) for 20 min or biotinylated anti-Ly6a (1:200) for 20 min followed by streptavidin conjugated Alexa Fluor 488 (1:1000; Life Technologies Cat#

S11223; RRID: AB_2336881) for 20 min prior to flow cytometry. Sca1+ cells were collected for Drop-seq.

Sertoli cell enrichment: Sertoli cells were enriched by selection of Amh+ or Sox9+ cells. For Amh+ cell selection, 129S.FVB-Tg(Amh-cre)8815Reb/J (JAX@mice, stock no. 007915) mice were crossed with *Rosa^{mT/mG}*. *Amh-cre*; *Rosa^{mT/mG}* were used to collect all live Amh positive (GFP expressing) and tdTomato negative cells by flow cytometry. For Sox9+ cell selection, Sox9-eGFP (MGI ID: 3844824) mice were used to collect GFP+ cells by flow cytometry.

Drop-seq procedure: Single-cell suspensions were diluted to 280 cells/ml and processed as described previously (Macosko et al., 2015). Briefly, cells, barcoded microparticle beads (MACOSKO-2011–10, Lots 113015B and 090316, ChemGenes Corporation), and lysis buffer were co-flow into a microfluidic device and captured in nanoliter-sized droplets. After droplet collection and breakage, the beads were washed, and cDNA synthesis occurred on the bead using Maxima H-minus RT (Thermo Fisher Scientific) and the Template Switch Oligo (Table S7). Excess oligos were removed by exonuclease I digestion. cDNA amplification was done for 15 cycles from a pool of 2,000 beads using HotStart ReadyMix (Kapa Biosystems) and the SMART PCR primer (Table S7). Individual PCRs were purified and pooled for library generation. A total of 600 pg of amplified cDNA was used for a Nextera XT library preparation (Illumina) with the New-P5-SMART PCR hybrid oligo, and a modified P7 Nextera oligo with 10 bp barcodes (Table S7). Sequencing was performed on a HiSeq-2500 (Illumina) in Rapid mode for read length of 112 nt, 115 nt, 126 nt, or 151 nt with the Read1CustomSeqB primer (Table S7). Oligo sequences are the same as previously described (Macosko et al., 2015).

Validation of ILCII cell population: Testes were collected from adult C57BL/6 (JAX@mice, stock no. 000664) mice and enzymatically and mechanically dissociated into a single cell suspension enriched for interstitial cells as described above. The single cell suspension is stained with a series of cell surface markers including anti-CD8-BV570 (1:300; Biolegend 100739), anti-CD3-PE/Cy7 (1:300; Biolegend 100219), anti-CD4-BV510 (1:300; Biolegend 100449), anti-IL7r-PE/Cy5 (1:300; Biolegend 135015). However, for intracellular staining - the ILCII cells were first incubated with PMA (10 ng / mL), ionomycin (10 μ M), with the addition of Golgi-stop reagent (BD Bioscience, San Jose, CA) for 4 h at 37°C and subsequently stained for anti-IL4-PE (1:300; Biolegend 504103), anti-IL-13-APC (1:300; Novus 011818), and anti-Gata3-BV421 (1:300; Biolegend 653813) for 30 minutes prior to flow cytometry.

Validation of Tcf21+ cells: 8week Tcf21-creERT2:tdtomato males were injected with a single dose of 2mg. The testes were collected within 24 hours, embedded in OCT, sectioned and stained with DAPI.

Histological Methods

Validation by single molecule RNA FISH: Adult male mice were perfused with 4% RNase-free paraformaldehyde (PFA). Testes were transferred to 20% RNase-free sucrose

overnight at 4°C, embedded in OCT, and cryosectioned at a 7µm thickness. Single molecule RNA FISH (smFISH) was performed as previously described by Raj et al. (Raj et al., 2008) with slight modifications. Briefly, tissue sections were warmed to room temperature for 10 min, cross-linked with 4% RNase-free PFA for 10 min, washed with 2x SSC, and permeabilized in 70% ethanol overnight at 4°C. Coverslips were pre-equilibrated in FISH wash buffer (2x SSC, 10% deionized formamide (Fisher Scientific)) for 5 min, and washed twice in FISH wash buffer containing 0.2% Tween-20 (Sigma) for 15 min for additional permeabilization. Custom probe sets targeting mouse *Gas6*, *Mfge8*, *Drd4*, *Dpysl4*, *Esyt3*, *Zfp3611*, *P2rx2*, and *Prm2* were labeled with Quasar 570 or Quasar 670 from Stellaris (LGC Bioscience Technologies). For probe design see Raj and Tyagi 2010 (Raj and Tyagi, 2010). Probes were added to hybridization buffer (2x SSC, 10% dextran sulfate, 1 µg/µl tRNA (Roche), 2 mM vanadyl ribonucleoside complex (NEB), 0.5% RNase-free BSA (Ambion), 10% deionized formamide) at 20 to 100 nM and applied to the tissue sections for 20 hrs at 37°C in a humidified chamber. Samples were washed with FISH wash buffer at room temperature for 15 min, at 37°C for 20 min, and again at 37°C with 20 ng/ml DAPI for 30 min, at room temperature for 10 min, and mounted using an in-house antifade solution (2 mg/ml p-phenylenediamine (Sigma) dissolved first in 0.3 M Tris-pH 9.0 and then brought up in glycerol at a ratio of 3:7). Tubules were imaged at 100x using an oil immersion objective on a Nikon Ti-E inverted fluorescence microscope equipped with a Photometrics Prime 95B Back-illuminated sCMOS Camera (Nikon Instruments Inc., Melville, NY). Each tubule was captured with 4 Z-stack images, each comprising 19 Z-sections with 0.3 µm spacing, and stitched together using NIS elements software. Quantification of spots was conducted using MATLAB (MathWorks) image analysis software developed and kindly provided by Arjun Raj. smFISH molecules were counted from maximum projection images for each whole tubule section after background subtraction of auto-fluorescent spots in the 488-nm channel. To determine the stage of the seminiferous epithelial cycle, we added PNA lectin (Life Technologies) at 1:650 and incubated tissue for 45 min. Each tubule was then reimaged, capturing Z-stacks as described above, and assigned to stages of spermatogenesis as described by Meistrich ML and Hess RA (Meistrich and Hess, 2013).

Validation by single molecule RNA HCR: Testes sections from adult male mice were prepared as described above, and single molecule RNA HCR was performed as described by Choi et al. (Choi et al., 2018) with slight modifications. Briefly, tissue sections were warmed to room temperature for 10 min, cross-linked with 4% RNase-free PFA for 10 min, washed with 1X PBS, and permeabilized in 0.5% Triton X-100 (Sigma T8787) for 1 hour at room temperature. Coverslips were then washed in 5X SSCT (Life Technologies 15557-044) and equilibrated in HCR hybridization buffer (30% Formamide (Ambion AM9342), 5X SSC (Life Technologies 15557-044), 9 mM Citric acid (pH 6.0) (Sigma 791725), 0.1% Tween-20 (Life Technologies 00-3005), 50 µg/mL Heparin (Sigma H3393), 1X Denhardt's solution (Life Technologies 750018), and 10% Dextran Sulfate (Sigma D6001)) for 1 hour at 37°C, then probes were added to the sections at a final concentration of 2nM and hybridized overnight in a humidified chamber at 37°C. Custom probe sets for *Esyt3*, *Lgals1*, *Mical2*, *Ptpvr*, and *Caskin1* were designed and synthesized by Molecular Technologies. Samples were washed three times in HCR wash buffer (30% Formamide (Ambion AM9342), 5X SSC (Life Technologies 15557-044), 9 mM Citric acid (pH 6.0) (Sigma 791725), 0.1% Tween-20

(Life Technologies 00–3005), 50 µg/mL Heparin (Sigma H3393)) for 15min at 37 °c and then three times for 15min at 37 °c in 5X SSCT (Life Technologies 15557–044). The probe sets were amplified with HCR hairpins for 45–90min at room temperature in HCR amplification buffer (5X SSC (Life Technologies 15557–044), 0.1% Tween-20 (Life Technologies 00–3005), 10% Dextran Sulfate (Sigma D6001), and 100ug/mL salmon sperm DNA (Thermo Fischer 15632–011)). Fluorescently-conjugated DNA hairpins used in the amplification were ordered from Molecular Technologies. Prior to use, the hairpins were ‘snap cooled’ by heating at 95°C for 90 seconds, and letting cool to room temperature for 30 min in the dark. After amplification, the samples were washed in 5X SSCT (Life Technologies 15557–044) and stained with 20ng/mL DAPI before being mounted with the in-house antifade solution previously described. Microscopy and quantification of single-molecule RNA spots were performed as described above. To achieve sequential hybridizations, HCR probes and hairpins were stripped from the samples by a 4-hour incubation with DNase I (Thermo Fischer 18068–015) at room temperature. Samples were then washed in 5X SSCT (Life Technologies 15557–044) before being hybridized with the next round of probes. RNA integrity of the samples was confirmed by the multispectral overlap between two probe sets designed against Pkg1, as described in Shah et. al. (Shah et al., 2016). Images from sequential hybridizations were aligned using fluorescent nanodiamonds (AdamasNano NDNV100nmHi10ml).

Computational Methods for Drop-seq Data

Preprocessing of Drop-seq data

Read filtering and alignment.: Raw paired-end sequence data were converted to queryname-sorted BAM files using *Picard* v2.6.0 FastqToSam (Broad Institute, 2016), and processed using *Drop-seq tools* v1.12 from the McCarroll laboratory as described previously (Macosko et al., 2015; Shekhar et al., 2016). Briefly, the first read is comprised of, from left to right, a 12-base cell-barcode, an 8-base unique molecular index (UMI), and a poly-T segment with 6 bases or longer. Read pairs with a base quality of less than 10 for any base of the cell barcode were removed. The second read (with read length of 112 nt, 115 nt, 126 nt, or 151 nt for different batches) was trimmed at the 5’ end to remove any SMART adaptor sequence, and at the 3’ end to remove poly-A tails of 6 consecutive bases or greater. The trimmed reads were then aligned to either the mouse reference genome (GRCm38, version 38) or a combined mouse (GRCm38) - human (GRCh37) mega-reference using *STAR* v2.5.2b (Dobin et al., 2013) with default settings. Reads uniquely mapped to the sense strand of gene exons were recorded and grouped by cell barcode. Throughout this study we used the Ensembl transcriptomic annotation (GRCm38 from Ensembl release 81, GRCh37 from Ensembl release 75, small RNA annotation from miRBase release 21, July 17, 2015).

Correcting for barcode synthesis errors.: The ChemGenes beads we ordered contained ~5% –10% cell barcodes that are identical in the first 11 bases, and with >95% of “T” at the last position of the UMI, as having been noticed before (Shekhar et al., 2016). These beads were missing a single base of cell barcode (Shekhar et al., 2016). Thus, the 20-bp barcode read would be expected to contain a mixed base at position 12 (the first base of the UMI) and a fixed T at position 20 (the first base of the polyT segment). To correct for this, we used *Drop-seq tools* DetectBeadSynthesisErrors to identify cell barcodes with mixed bases at

position 12 and UMIs with fixed “T” at position 8 and to insert an “N” at cell barcode position 12 for these reads with synthesis missing base. This resulted in a corrected set of cell barcodes and UMIs that was used for the estimation of digital gene expression. 5%–10% of cell barcodes were corrected in this way in all the datasets.

Digital gene expression.: To distinguish cell barcodes that represent genuine transcriptomic libraries arising from cells, rather than from empty droplets, we ordered the cell barcodes by the total number of reads per cell barcode and estimated the inflection point in the cumulative reads distribution plot, as described previously (Macosko et al., 2015). All cell barcodes with the total number of transcripts larger than this cutoff were extracted for downstream analysis.

To digitally count gene transcripts, UMIs in each gene and within each cell were assembled, and UMIs within edit distance of 1 (substitutions only) were collapsed. The total number of unique UMI sequences was counted and reported as the transcript count for that particular gene in a given cell. This resulted in a digital gene expression matrix (DGE) with genes as rows and cells as columns that served as the starting point for downstream analyses.

Filtering for cells and genes.: The starting pool of 59,313 cells and 37,241 genes was first selected by the cell size and integrity filters – cells with ≥ 500 detected genes per cell or with $\geq 10\%$ of transcripts corresponding to mitochondria-encoded genes were removed. We then removed low abundance genes that were detected in ≤ 15 cells or with ≤ 20 UMIs summed across all the retained cells. These filters resulted in 34,633 cells and 24,947 genes, which were considered for further analysis. The information of batch origin, number of retained cells, the total number of UMIs (nUMI) and detected genes (nGene) for each dataset was summarized in Table S1.

Among the retained cells, the average number of detected genes per cell was 2,057 (IQR 864 – 3,006, Figure 1D) and the average number of UMIs was 6,205 (IQR 1,435 – 8,074). The average number of transcriptome-mapped reads per cell was 18,699 (IQR 4,124 – 23,627). The mean number of reads for a given UMI was 3.3 (IQR 2.2 – 3.3). On average, 95% of detected genes has a transcript count of ≤ 10 (54.6% 1’s, 17.8% 2’s, 8.5% 3’s, 4.9% 4’s, 3.2% 5’s and $<2.2\%$ each for 6’s – 10’s).

For each cell, we normalized transcript counts by (1) dividing by the total number of UMIs per cell and (2) multiplying by 10,000 to obtain a transcripts-per-10K measure, and then log-transformed it by $E = \ln(\text{transcripts-per-10K} + 1)$. For PCA, we used standardized expression values obtained by centering and scaling for each gene using $(E - \text{mean}(E)) / \text{sd}(E)$.

Evaluation of technical variability

Between-batch reproducibility.: We analyzed reproducibility across the six batches of 7- to 9- week old wild-type male mice seminiferous tubule (ST) data, which were the unbiased representation of ST. All 6 ST batches contained $\sim 2\text{k}$ cells passing cell size and integrity filters and $\sim 19\text{k}$ genes passing abundance filter in the remaining cells. We performed PCA on each of the six batches. The placement of the five major cell types (somatic cells and the four recognized major germ cell types) in PC1-PC2 plots were similar across the six batches,

indicating reproducible patterns of heterogeneity (Figure S1B). We used scree plot and Jackstraw permutation to identify top PCs that explain statistically significant proportion of the variance among the genes. We performed Louvain-Jaccard clustering using the top 13 significant PCs for each of the six batches and obtained 14 clusters for each. We then ordered clusters within each batch using custom scripts. Briefly, for each batch, we calculated cluster centroids of each gene as $\ln(\text{mean of normalized expression} + 1)$ over all cells in each cluster, and obtained Euclidean distances for the cluster centroids, which were then ordered using the optimal leaf ordering algorithm in R Package *Seriation*. The cluster IDs were thereby renumbered according to the seriation. To compare the cluster solutions among the six batches, we cross-tabulated rank correlation coefficients among all pairs of cluster centroids across the six ST batches, which demonstrated that clusters were largely reproducible across batches (Figure S1C).

Evaluating targeted depletion or enrichment experiments.: We performed PCA on the entire expression matrix of 34,633 cells and 24,947 genes and visualized subset depletion or enrichment by the cell density differences in the PC1-PC2 space. Specifically, density of cell counts (d) was calculated in PC1-PC2 space in a 50–50 grid for each of the five experiments using *hist2d* function in R and then log-transformed as $c = \ln(d+1)$ (Figure S1D, *top panel*). The per-grid enrichment/depletion pattern of each depletion/enrichment experiment against the original ST experiment ($n = 6$ ST batches, c_0) was calculated as the ratio $r = c / \sum(c) - c_0 / \sum(c_0)$, where c is the per-grid count of cells on the log scale. The density ratio against the original ST experiment (Figure S1D, *bottom panel*) confirmed that our targeted-subset enrichment experiments have indeed enriched for the intended rare cell types: the 1n-depleted ST datasets showed depletion of spermatids; the SPG-enriched datasets had enrichment of spermatogonia cells; the INT-enriched and SER-enriched datasets had enrichment of the interstitial cells and the Sertoli cells, respectively.

The overall atlas for spermatogenesis

Dimensionality reduction using PCA.: We performed principal component analysis (PCA) on the standardized gene expression matrix with 34,633 cells and 24,947 genes using the PCA function in R package *Seurat v1.4.0.3* (Satija et al., 2015). We used elbow point in the scree plot and the distribution of eigenvalues to identify top PCs and chose the top 55 PCs for downstream clustering. PCA thus reduced the dimensionality of expression data from 24,947 (# of genes) to 55 (# of selected PCs).

Louvain-Jaccard clustering for 24 ST batches.: We initially performed PCA on 24 ST batches with 33,180 cells passing the cell size and integrity filter and 24,482 genes passing the abundance filter in the remaining cells. These 24 ST batches did not include INT6 batch. We did Louvain-Jaccard clustering using top 40 PCs. As an initial attempt to assess cellular heterogeneity, we used the FindClusters function in the R package *Seurat* to identify cell clusters. This approach used the top PC scores to calculate Euclidean distance among all pairs of cells and for each, identified its 30 nearest neighbors. It then used the Louvain method of clustering using the Jaccard distance among cells as weights, where the Jaccard distance between any two cells was defined as the degree of overlap of their 30 nearest neighbors. The Louvain method for community detection (Blondel et al., 2008) is a greedy

optimization method that runs in time $O(n \log n)$ (n is the number of cells). For the 33,180 cells from the first 24 ST batches, we initially identified 31 clusters, which were merged as described below.

Merging clusters and annotation to known cell types.: Our first step is to identify major cell types in the overall atlas, and we will rely on focused re-clustering in the second step to identify subtypes within each major cell type. To achieve the first goal, we merged some clusters in the initial set of 31 by an iterative process that combines statistical evaluation and biological assignment. We computed a 31–31 Euclidean distance matrix for the 31 cluster centroids, which were defined as $\ln(\text{mean of normalized expression} + 1)$ over all cells in each cluster. We ordered and renumbered the 31 clusters using the optimal leaf ordering algorithm in the R Package *Seriation*. We selected marker genes for each cluster by comparing the expression level in one cluster against that in all other clusters using a nonparametric binomial test. The marker gene selection criteria include: (1) at least a 20% difference in detection rate; (2) a minimum of 2-fold higher mean expression level in the cell type compared to all other cell types, and (3) $p\text{-value} < 0.01$ in the binomial test. Pairs of neighboring clusters without a single differentially-expressed gene were merged. Further merging of the ordered clusters was based on known markers for major cell types. Overall, for the 31 ordered clusters, we assigned 7 neighboring clusters as somatic cell group, 3 neighboring clusters as SPG, 2 neighboring clusters as two transitioning populations between SPG and Scytes, 7 neighboring clusters as Scytes, 8 neighboring clusters as round spermatids, and 4 neighboring clusters as elongating spermatids.

Louvain-Jaccard clustering for 25 ST batches.: We then included an additional batch, INT6, with only somatic cells, into the analysis. For the 25 batches after adding INT6, we performed Louvain-Jaccard clustering using top 55 PCs, and obtained 10 clusters using minimal resolution. We selected markers for each of these 10 clusters using the same method as above and then based on the markers, merged these 10 clusters into 2 major groups – 1 somatic cell group and 1 germ cell group, which confirmed that INT6 only contributes to somatic cell group. Since INT6 did not contribute to germ cell group, we retained the major cluster assignment for the germ cells from the 24 ST batches as described above. We then performed focused clustering on the somatic cell group (5,081 cells) and obtained 7 somatic cell types as described in details in *Focused analysis 2: the somatic cells* section. As a result, the 25 batches led to the identification of 11 major cell types: four germ cell types (spermatogonia – including 1 SPG cluster and 2 transitioning clusters between SPG and Scytes, spermatocytes, round spermatids and elongated spermatids) as well as seven somatic cell types (Sertoli, a Mesenchymal cell population, an ILCII population, Macrophage, Endothelial, Myoid, and Leydig cells).

Cellular attributes.: We also calculated a series of per-cell attributes based on the transcriptome data, and included them in our GEO submission **GSE112393** and shown in Figure 1D.

- %Mito, is the percent of UMIs accounted for by the 24 mitochondria-encoded genes. It serves as an index of cell injury or viability, with a smaller %Mito indicating a healthier cell.

- %ChrX, %ChrY, are the percentage of transcripts accounted for by the 2,062 ChrX or 643 ChrY Ensembl genes, respectively.
- Gini index, is a measure of gene expression inequality, using either all genes or only the detected genes
- Cell cycle indices. We obtained a list of 590 cell cycle genes, and its partition into five phase-specific lists: G1S (N=98), S (N=106), G2M (N=130), M (N=151), and MG1 (N=105) (Macosko et al., 2015). The cell cycle index (%cell cycle) is calculated as the proportion of all 590 cell cycle genes in the total UMI count of the cell. The cell phase index, one for each phase, is the relative proportion for one phase-specific list of genes over the UMI counts of all cell cycle genes.

Marker selection for the 11 major cell types and pathway analysis.: Markers for each of the 11 major cell types were obtained by comparing a given cell type with all other 10 cell types using the binomial likelihood test embedded in R package *Seurat v1.4.0.3* (results are in Table S2). Selection criteria are (1) at least 20% difference in detection rate; (2) a minimum of 2-fold higher mean expression level in the cell type compared to all other cell types, and (3) p-value < 0.01 in the binomial test. To display the markers in the 11-centroid heatmap while accommodating their wide range of absolute expression levels, we centered each marker's expression levels across the 11 centroids and scaled by its standard deviation (Figure 1C).

For functional enrichment analysis of the 11 marker lists we use the Gene Ontology terms and *GOrilla* (Eden et al., 2009) (accessed on Aug 14, 2017) using the 37,241 Ensembl detected genes as the background list. GO terms that contain too many (>500) or too few (<50) genes were removed as their enrichment p-values were either too easy to be significant or too unstable (Figure 1C).

Focused analysis 1: developmental trajectory of germ cells

Focused analysis for germ cells.: After evaluating cellular heterogeneity in the global atlas (Figure 1A), we performed iterative, zoomed-in analyses to assess subtypes contained within the major cell types. As germ cells tend to have higher number of genes and unique UMIs compared to somatic cells (Figure 1D), we raised the cell size filter and only selected germ cells with >1k detected genes for downstream analysis. We extracted normalized expression for germ cells (20,646 cells with >1k detected genes, 24,475 non-0 genes). We standardized for each gene by centering and scaling and selected 2,047 highly variable genes (HVG) for germ cells using MeanVarPlot function in R package *Seurat*. We performed focused PCA on germ cells with >1k detected genes (N=20,646) using the 2,047 HVG and calculated Jaccard distance using top 24 PCs. We visualized the expression levels of known germ cell markers in 2D PCA plot, with the darkest blue indicating no detected expression and the darkest red indicating highest expression for the marker (Figure 2B).

Focused clustering for non-SPG germ cells.: From focused analysis of germ cells (i.e., removing the identified somatic cells and smaller cells), we found that the SPG cells

(including 1 SPG cluster and 2 transitioning clusters between SPG and Scytes) are separated from the continuous transitions of other germ cells (Figure 2A). To closely examine the developmental trajectory of the germ cells after SPG, we extracted the normalized gene expression matrix for the 18,450 non-SPG germ cells, which contains 24,460 detected genes. We selected 1,879 HVG using the MeanVarPlot function in R package *Seurat*. For these non-SPG germ cells, we performed focused PCA using only HVG, and obtained 9 clusters by Louvain-Jaccard clustering using the top 20 PCs. Thereby, we obtained a total of 12 germ cell clusters, including 1 SPG cluster, 2 transitioning clusters between SPG and Scytes, and 9 non-SPG germ cell clusters.

Ordering of germ cell clusters by seriation.: To order the 12 germ cell clusters, we first computed an expression matrix with 12 cluster centroids and 24,475 genes by $\ln(\text{mean of normalized expression} + 1)$ over all cells in each cluster for each gene and obtained Euclidean distances for each pair of cluster centroids. We ordered the cluster IDs of Euclidean distance matrix using optimal leaf ordering (OLO) algorithm in R Package *Seriation v1.2.2* (Hahsler et al., 2008). The cluster IDs were then renumbered according to the seriation and neighboring clusters were thereby the most similar. We visualized rank correlation of the 12 ordered germ cell cluster centroids (Figure 2B), as well as cell-cell rank correlation and Jaccard distance of all 20,646 germ cells ordered by the 12 ordered germ cell clusters (Figure S2A) in heatmaps.

Comparison with pseudotemporal ordering by two other methods.: We also inferred pseudotemporal ordering by applying two methods widely adopted for this purpose: *Waterfall* and *Monocle*. *Waterfall* (Shin et al., 2015) contains three steps: pre-processing by removing outlier cells and determining route and orientation; reconstructing of a trajectory using minimum spanning tree and determining the pseudotime of individual cells; identifying genes correlated with the pseudotime. We followed the standard pipeline to calculate pseudotime for 20,238 germ cells using 9,074 genes retained by *Waterfall*. The results are in good agreement with the cells' assignment to our 12 germ cell cluster (Figure S2C, left panel). We showed that gene expression dynamics for 8 representative known germ cell markers predicted by *Waterfall* agree well with our germ cell clusters (Figure S2D). *Monocle 2* (Qiu et al., 2017; Trapnell et al., 2014) performs clustering and selects differentially-expressed genes from dense cell clusters for use in trajectory reconstruction. We applied dpFeature in *Monocle 2* to select highly informative genes which 1) excludes genes expressed in <5% of cells; 2) performs PCA using prcomp_irlba; 3) projects cells to 2D using tSNE; 4) clustered by densityPeak; 5) performed differential gene expression analysis and selected the top 1000 significant differentially-expressed genes. We then reduced dimension with DDRTree and ordered cells along the trajectory using *Monocle 2*. The result was compared with that from the 12 germ cell clusters (Figure S2C, right panel).

Comparison with ordering of non-SPG germ cells by SOM.: We performed clustering for non-SPG germ cells (18,450 cells with >1k detected genes, 1,879 HVG) using unbiased self-organization map (SOM) embedded in R package *som v0.3.5.1* (Yan, 2016) and obtained 20 ordered clusters for non-SPG germ cells. We then visualized SOM 20 clusters in 2D PCA view with alternating colors for neighboring clusters (Figure S2F, left panel), which agreed

well with our ordering of the 12 germ cell clusters. As the cells describe a long “slender” trajectory, we were not only able to identify 20 clusters but also found that they were robust. That is, neighboring clusters can be distinguished by highly informative markers. After filtering with both fold-change and p-values we identified cluster-specific markers that are often expressed highly in a given cluster but substantially lower in its flanking clusters. The heatmap for 508 markers (Figure S2F, right panel) account for the distinction among the 20 clusters with 14–44 transient transcripts for each cluster. The color scale from white-to-red spans the range of 0–4.2 in the natural log of cluster centroid expression levels.

Marker selection for the 12 germ cell clusters.: We obtained markers for the 12 germ cell clusters (20,646 cells) by comparing each cluster against all other 11 germ cell clusters using binomial likelihood test in Seurat *v1.4.0.3* (Table S3A). As before, the criteria are: (1) At least 20% difference in detection rate; (2) a minimum of 2-fold difference in mean expression level in the cluster-in-question compared to all other clusters, and (3) p-value < 0.01 in the binomial test.

Discovery of co-regulated gene groups using the 12 germ cell cluster centroids.: To identify the genes exhibiting concerted dynamic patterns along the course of spermatogenesis, we performed unbiased gene clustering by SOM using the centroids of the twelve germ cell clusters. We focused on 8,583 highly expressed (mean>2) and highly variable (variance/mean>0.5), where the mean and variance were calculated over the 12 cluster centroids, each of which was defined as the median vector of all the cells in each cluster. In the 10–10 configuration, we used SOM to separate the genes into 100 groups, linked in a 10–10 grid, resulting in 5 to 729 genes in each group (median: 35.5 genes). Re-running SOM with a 6–6 grid led to similar results (not shown). Gene groups in the four corners of this “map” showed distinct expression patterns, e.g., those in the upper left and lower left had highest expression in spermatogonia and round spermatids, respectively, whereas those in the upper right and lower right showed highest expression in spermatocytes and elongated spermatids, respectively. In an alternative approach, we applied unsupervised k-means clustering to identify major gene groups among the 8,583 genes at k=6 (Figure 3A) and k=12 (Figure S3A). At k=6, the six gene groups map to the four corners of the 10–10 SOM grid, containing genes with highest expression during distinct phases of the spermatogonia-to-spermatocyte transition (Figure S3B). The centroid expression levels of the 8,583 genes and their membership in the SOM and k-means clustering were included in Table S3B serving as a reference resource of the dynamic gene regulation in germ cell development.

Further, to uncover potential transcription factors driving the dynamic expression patterns we applied motif enrichment analyses for the sequences within \pm 1kb of the transcription start site for genes within each of the six gene groups (Bailey et al., 2009; Machanick and Bailey, 2011) (Figure 3B).

Expression patterns of mammalian infertility genes in the 12 germ cell cluster centroids.: We extracted from Matzuk et al. (Matzuk and Lamb, 2008) >200 genes exhibiting fertility phenotypes when carrying loss-of-function mutations. The genes that were indicated as *infertile* by Matzuk et al. were combined with those indicated as *secondary infertility*,

selective infertility, and *mostly infertile*. Genes resulting only in female infertility were excluded. Among 209 such male infertility genes, 192 were detected in the germ cell populations in our study. To curate male infertility genes in humans, we queried the Online Inheritance in Man (OMIM) database (McKusick-Nathans Institute of Genetic Medicine, 2017) and identified 260 genes involved in heritable and often rare cases of infertility (accessed on 10/30/2017). To illustrate the utility of the reference map of dynamic expression, we applied hierarchical clustering to the 114 mouse infertility genes and the 87 human infertility genes meeting the same filtering criteria that led to the 8,583 high-variability genes, and plotted expression heatmap for these mouse and human genes across 12 germ cell cluster centroids (Figure 3C).

Further, based on literature review of previous loss-of-function studies, we categorized 150 of the detected mouse infertility genes according to their earliest observed defects in major cell types during spermatogenesis: spermatogonia, spermatocyte, round spermatid, elongated spermatid, or sperm. The expression heatmaps for these five subsets of genes were shown in Figure 3D

Further focused analysis of the spermatogonia cells

Focused subset clustering and ordering of SPG cells.: In order to elucidate the heterogeneity within the SPG cluster, we performed focused PCA for only SPG cells in the SPG cluster: 2,484 cells with >1,000 UMIs, and 21,541 detected genes. We obtained four subtypes of SPG cells using the Louvain-Jaccard clustering with the top 10 PCs, and visualized them using tSNE (Figure 4A). In order to order the 4 SPG subtypes, we first calculated the SPG subtype centroids as $\ln(\text{mean of normalized expression} + 1)$ across all cells in each subtype for each gene, and then obtained pairwise Euclidean distances for the 4 SPG subtypes and ordered them using the optimal leaf ordering (OLO) algorithm in *R* Package *Seriation*. The cluster IDs were henceforth numbered according to the ordering.

Marker selection for 4 SPG subtypes.: Markers for each SPG subtype were obtained by comparing cells in each cluster against those in the other 3 clusters, using the binomial likelihood test in Seurat *v1.4.0.3*. Genes with a minimum 1.5-fold effect size and $p < 0.01$ were selected (Figure 4B, Table S5A). We visualized per-cell expression patterns of representative markers in tSNE plot (Figure 4C). We summarized the schematic depicting the position of 4 SPG subtypes across stages of seminiferous epithelial cycle inferred from the expression patterns of markers for each SPG subtype (Figure 4D).

Heterogeneity within undifferentiated SPG cells (SPG1 subtype).: The SPG subtype 1 (SPG1) is undifferentiated based on known stem cell marker genes. We analyzed heterogeneity within the 213 SPG1 cells by re-clustering cells with >1k UMIs by three methods, using: 1) all detected genes in the SPG1 subtype (n=15,765); 2) highly variable genes (HVG) (n=888); and 3) known spermatogonial stem cell markers (n=32). Clustering was performed using top 3 significant PCs for each of the three gene sets. Cross-tabulation of the number of cells for each cluster across clustering solutions with the 3 sets of genes (right panel) showed inconsistency among the three sets of clusters.

Focused analysis 2: the somatic cells—In order to assess cellular heterogeneity among the somatic cells (Figure 1B), we performed focused PCA on the 5,081 somatic cells, using 22,734 genes detected in >3 cells (Figure 1B **insert**). We performed Louvain-Jaccard clustering using the top 23 significant PCs and identified seven somatic clusters (Figure 5A). They correspond to five known cell types (based on known markers, Figure 5B), as well as two previously unknown cell populations. One of them was subsequently identified as the innate lymphoid type II cells. Heatmap for Jaccard distance of all 5,081 somatic cells ordered by the 7 clusters showed that these 7 clusters are distinct from each other (Figure S1E).

Distinct expression profiles of transcription factors—A list of 1,675 mouse transcription factors was downloaded from the non-redundant mouse set of the *Riken Transcription Factor Database (TFdb)* on Oct 4, 2017 (Table S4A). We performed three sets of analysis for differentially-expressed TFs, 1) across 12 germ cell stages; 2) across 4 SPG subtypes; 3) across 7 somatic cell types (Table S4). For each set of analysis, differentially expressed TFs were identified by comparing each cluster against all other clusters using the binomial likelihood test in *Seurat v1.4.0.3*. The criteria are: (1) at least 20% difference in detection rate; (2) a minimum of 2-fold higher mean expression level in the cluster-in-question compared to all other clusters, and (3) p-value < 0.01 in the binomial test.

Among the 1,097 TFs expressed in the 20,646 germ cells, 110 were differentially expressed across the 12 germ cell clusters (Table S4B). Among the 1,065 TFs present in the 2,484 SPG cells, 57 were differentially expressed across the 4 SPG subtypes (Table S4C). Among the 1,025 TFs present in the 5,081 somatic cells, 92 were differentially expressed across the 7 somatic cell types (Table S4D). For each TF, the cluster centroids were calculated as described above. We visualized per-cell expression patterns of representative differentially-expressed TFs for 4 SPG subtypes in tSNE plot (Figure S4B).

Further focused analysis of Sertoli cells

Focused clustering of Sertoli cells.: To assess cellular heterogeneity among the Sertoli cells, we performed PCA on 1,067 Sertoli cells with >1,000 detected genes. We obtained nine clusters by Louvain-Jaccard clustering using the top 11 PCs (Figure 6B). Marker selection was performed as before. A k-means clustering analysis at k=4 found that the nine clusters can be grouped into four main clusters. To capture this nested clustering pattern we named the nine clusters as 1, 2A-2B, 3A-3B, and 4A-B-C-D.

Comparing nine functional clusters with four known stages of Sertoli cells in the seminiferous tubule cycle.: Mouse genes with Sertoli cell stage-specific expression patterns were retrieved from previous studies (Hasegawa and Saga, 2012; Wright et al., 2003) and divided into four groups, representing four stage groups of the seminiferous epithelial cycle: stages I-III, IV-VI, VII-VIII, and IX-XII. These four lists of genes were used to calculate the relative expression level of stage-specific genes, defined as a percentage of one stage over all 4 stages (akin to the earlier analysis of cell cycle fractions for G1S, S, G2M, M, MG1). The four-stage fractions were calculated for each of the 9 molecularly defined clusters of Sertoli cells (Figure 6C).

QUANTIFICATION AND STATISTICAL ANALYSIS

For the Drop-seq experiments, cells were collected from 25 experiments, including 6 experiments with unbiased representation of mouse seminiferous tubules, 2 experiments with 1n-depletion, 3 experiments with spermatogonia-targeted enrichment, 5 experiments with interstitial-targeted enrichment, and 8 experiments with Sertoli-targeted enrichment (See Table S1).

For smFISH experiments, the numbers of seminiferous tubules imaged from one mouse are indicated in Figures S7C-D. For immunofluorescence imaging, representative images were chosen from 40 images of one mouse. Precision measures are listed in the appropriate figure legends (Figure 5F, mean \pm SD; Figures S7C-D, mean \pm SEM).

Statistical methodologies and software packages used are described according to the STAR Methods format. All analyses were performed in R.

DATA AND SOFTWARE AVAILABILITY

Data Resources—Raw and processed data files for Drop-seq experiments are available under the GEO accession number GEO: GSE112393.

R Markdown Code for Reproducing Clustering Analysis—As an accompaniment to this paper, we provide an R markdown file that describes step-by-step procedures, including the loading and preprocessing of the Drop-seq digital expression matrix, PCA, Louvain-Jaccard clustering, data visualization, ordering by seriation, and differential expression tests. The R commands are provided at https://github.com/qianqianshao/Drop-seq_ST.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank members of the Hammoud, Li, and Yamashita Lab for scientific discussions. We thank Dr. Shuichi Takayama and Priyan Weerappuli for synthesis of microfluidic devices (Department of Biomedical Engineering, University of Michigan). We thank Sethuramasundaram Pitchiaya for smFISH advice. We thank Drs. Yukiko Yamashita, Lei Lei and lab members for helpful comments on the manuscript. We thank Mary Ann Handel for great scientific discussions and feedback. We thank Shosei Yoshida for providing Gfra1 mice. This research was supported by National Institute of Health (NIH) grants 1R21HD090371-01A1 (S.S.H. and J.Z.L.) and 1DP2HD091949-01 (S.S.H.), and Michigan Institute for Data Science (MIDAS) grant for Health Sciences Challenge Award (J.Z.L. and S.S.H.), CTB training grant 5T32HD079342-04 (A.N.S.), MSTP training grant 5T32GM007863-38 (A.N.S.), and CMB training grant 2T32GM007315-41 (G.L.M.).

References

- Acharya A, Baek ST, Banfi S, Eskiocak B, and Tallquist MD (2011). Efficient inducible Cre-mediated recombination in Tcf21 cell lineages in the heart and kidney. *Genesis* 49, 870–877. [PubMed: 21432986]
- Aloisio GM, Nakada Y, Saatcioglu HD, Pena CG, Baker MD, Tarnawa ED, Mukherjee J, Manjunath H, Bugde A, Sengupta AL, et al. (2014). PAX7 expression defines germline stem cells in the adult testis. *The Journal of clinical investigation* 124, 3929–3944. [PubMed: 25133429]

- Anderson EL, Baltus AE, Roepers-Gajadien HL, Hassold TJ, de Rooij DG, van Pelt AM, and Page DC (2008). Stra8 and its inducer, retinoic acid, regulate meiotic initiation in both spermatogenesis and oogenesis in mice. *Proceedings of the National Academy of Sciences of the United States of America* 105, 14976–14980. [PubMed: 18799751]
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, and Noble WS (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic acids research* 37, W202–208. [PubMed: 19458158]
- Ball RL, Fujiwara Y, Sun F, Hu J, Hibbs MA, Handel MA, and Carter GW (2016). Regulatory complexity revealed by integrated cytological and RNA-seq analyses of meiotic substages in mouse spermatocytes. *BMC genomics* 17, 628. [PubMed: 27519264]
- Blondel VD, Guillaume J-L, RLambiotte R, and Lefebvre E (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics*, P10008.
- Braun RE, Behringer RR, Peschon JJ, Brinster RL, and Palmiter RD (1989). Genetically haploid spermatids are phenotypically diploid. *Nature* 337, 373–376. [PubMed: 2911388]
- BroadInstitute (2016). Picard.
- Buaas FW, Kirsh AL, Sharma M, McLean DJ, Morris JL, Griswold MD, de Rooij DG, and Braun RE (2004). Plzf is required in adult male germ cells for stem cell self-renewal. *Nature genetics* 36, 647–652. [PubMed: 15156142]
- Campbell JN, Macosko EZ, Fenselau H, Pers TH, Lyubetskaya A, Tenen D, Goldman M, Verstegen AM, Resch JM, McCarroll SA, et al. (2017). A molecular census of arcuate hypothalamus and median eminence cell types. *Nat Neurosci* 20, 484–496. [PubMed: 28166221]
- Chan F, Oatley MJ, Kaucher AV, Yang QE, Bieberich CJ, Shashikant CS, and Oatley JM (2014). Functional and molecular features of the Id4+ germline stem cell population in mouse testes. *Genes & development* 28, 1351–1362. [PubMed: 24939937]
- Chen C, Ouyang W, Grigura V, Zhou Q, Carnes K, Lim H, Zhao GQ, Arber S, Kurpios N, Murphy TL, et al. (2005). ERM is required for transcriptional control of the spermatogonial stem cell niche. *Nature* 436, 1030–1034. [PubMed: 16107850]
- Chen LY, Willis WD, and Eddy EM (2016). Targeting the Gdnf Gene in peritubular myoid cells disrupts undifferentiated spermatogonial cell development. *Proc Natl Acad Sci U S A* 113, 1829–1834. [PubMed: 26831079]
- Clermont Y (1972). Kinetics of spermatogenesis in mammals: seminiferous epithelium cycle and spermatogonial renewal. *Physiological reviews* 52, 198–236. [PubMed: 4621362]
- Costoya JA, Hobbs RM, Barna M, Cattoretti G, Manova K, Sukhwani M, Orwig KE, Wolgemuth DJ, and Pandolfi PP (2004). Essential role of Plzf in maintenance of spermatogonial stem cells. *Nat Genet* 36, 653–659. [PubMed: 15156143]
- Cui S, Ross A, Stallings N, Parker KL, Capel B, and Quaggin SE (2004). Disrupted gonadogenesis and male-to-female sex reversal in Pod1 knockout mice. *Development (Cambridge, England)* 131, 4095–4105.
- Danielian PS, Hess RA, and Lees JA (2016). E2f4 and E2f5 are essential for the development of the male reproductive system. *Cell cycle (Georgetown, Tex)* 15, 250–260.
- de Rooij DG (1998). Stem cells in the testis. *International journal of experimental pathology* 79, 67–80. [PubMed: 9709376]
- de Rooij DG, and Russell LD (2000). All you wanted to know about spermatogonia but were afraid to ask. *Journal of andrology* 21, 776–798. [PubMed: 11105904]
- DeFalco T, Potter SJ, Williams AV, Waller B, Kan MJ, and Capel B (2015). Macrophages Contribute to the Spermatogonial Niche in the Adult Testis. *Cell reports* 12, 1107–1119. [PubMed: 26257171]
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. [PubMed: 23104886]
- Eddy EM (2002). Male germ cell gene expression. *Recent Prog Horm Res* 57, 103–128. [PubMed: 12017539]
- Eden E, Navon R, Steinfeld I, Lipson D, and Yakhini Z (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10, 48. [PubMed: 19192299]

- El-Darwish KS, Parvinen M, and Toppari J (2006). Differential expression of members of the E2F family of transcription factors in rodent testes. *Reproductive biology and endocrinology : RB&E* 4, 63. [PubMed: 17147820]
- Endo T, Romer KA, Anderson EL, Baltus AE, de Rooij DG, and Page DC (2015). Periodic retinoic acid-STRA8 signaling intersects with periodic germ-cell competencies to regulate spermatogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 112, E2347–2356. [PubMed: 25902548]
- Evans E, Hogarth C, Mitchell D, and Griswold M (2014). Riding the spermatogenic wave: profiling gene expression within neonatal germ and sertoli cells during a synchronized initial wave of spermatogenesis in mice. *Biology of reproduction* 90, 108. [PubMed: 24719255]
- Fawcett DW, Ito S, and Slautterback D (1959). The occurrence of intercellular bridges in groups of cells exhibiting synchronous differentiation. *J Biophys Biochem Cytol* 5, 453–460. [PubMed: 13664686]
- Franca LR, Ogawa T, Avarbock MR, Brinster RL, and Russell LD (1998). Germ cell genotype controls cell cycle during spermatogenesis in the rat. *Biol Reprod* 59, 1371–1377. [PubMed: 9828180]
- Fukuda N, Fukuda T, Sinnamon J, Hernandez-Hernandez A, Izadi M, Raju CS, Czaplinski K, and Percipalle P (2013). The transacting factor CBF-A/Hnrnpab binds to the A2RE/RTS element of protamine 2 mRNA and contributes to its translational regulation during mouse spermatogenesis. *PLoS genetics* 9, e1003858. [PubMed: 24146628]
- Gaysinskaya V, Soh IY, van der Heijden GW, and Bortvin A (2014). Optimized flow cytometry isolation of murine spermatocytes. *Cytometry A* 85, 556–565. [PubMed: 24664803]
- Griswold MD (1995). Interactions between germ cells and Sertoli cells in the testis. *Biol Reprod* 52, 211–216. [PubMed: 7711190]
- Guo F, Li X, Liang D, Li T, Zhu P, Guo H, Wu X, Wen L, Gu TP, Hu B, et al. (2014). Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell stem cell* 15, 447–459. [PubMed: 25220291]
- Guo J, Grow EJ, Yi C, Mlcochova H, Maher GJ, Lindskog C, Murphy PJ, Wike CL, Carrell DT, Goriely A, et al. (2017). Chromatin and Single-Cell RNA-Seq Profiling Reveal Dynamic Signaling and Metabolic Transitions during Human Spermatogonial Stem Cell Development. *Cell stem cell* 21, 533–546 e536. [PubMed: 28985528]
- Hahsler M, Hornik K, and Buchta C (2008). Getting things in order: An introduction to the R package seriation. *Journal of Statistical Software* 25, 1–34.
- Hammoud SS, Low DH, Yi C, Carrell DT, Guccione E, and Cairns BR (2014). Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. *Cell Stem Cell* 15, 239–253. [PubMed: 24835570]
- Hammoud SS, Low DH, Yi C, Lee CL, Oatley JM, Payne CJ, Carrell DT, Guccione E, and Cairns BR (2015). Transcription and imprinting dynamics in developing postnatal male germline stem cells. *Genes & development* 29, 2312–2324. [PubMed: 26545815]
- Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, Saadatpour A, Zhou Z, Chen H, Ye F, et al. (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* 172, 1091–1107 e1017. [PubMed: 29474909]
- Handel MA (2004). The XY body: a specialized meiotic chromatin domain. *Experimental cell research* 296, 57–63. [PubMed: 15120994]
- Hara K, Nakagawa T, Enomoto H, Suzuki M, Yamamoto M, Simons BD, and Yoshida S (2014). Mouse spermatogenic stem cells continually interconvert between equipotent singly isolated and syncytial states. *Cell Stem Cell* 14, 658–672. [PubMed: 24792118]
- Hasegawa K, and Saga Y (2012). Retinoic acid signaling in Sertoli cells regulates organization of the blood-testis barrier through cyclical changes in gene expression. *Development (Cambridge, England)* 139, 4347–4355.
- Heimberg G, Bhatnagar R, El-Samad H, and Thomson M (2016). Low Dimensionality in Gene Expression Data Enables the Accurate Extraction of Transcriptional Programs from Shallow Sequencing. *Cell Syst* 2, 239–250. [PubMed: 27135536]

- Hermann BP, Mutoji KN, Velte EK, Ko D, Oatley JM, Geyer CB, and McCarrey JR (2015). Transcriptional and translational heterogeneity among neonatal mouse spermatogonia. *Biology of reproduction* 92, 54. [PubMed: 25568304]
- Hofmann MC, Braydich-Stolle L, and Dym M (2005). Isolation of male germ-line stem cells; influence of GDNF. *Dev Biol* 279, 114–124. [PubMed: 15708562]
- Hogarth CA, Evanoff R, Mitchell D, Kent T, Small C, Amory JK, and Griswold MD (2013). Turning a spermatogenic wave into a tsunami: synchronizing murine spermatogenesis using WIN 18,446. *Biol Reprod* 88, 40. [PubMed: 23284139]
- Hoyer-Fender S (2003). Molecular aspects of XY body formation. *Cytogenetic and genome research* 103, 245–255. [PubMed: 15051945]
- Inoue K, Ichiyanaagi K, Fukuda K, Glinka M, and Sasaki H (2017). Switching of dominant retrotransposon silencing strategies from posttranscriptional to transcriptional mechanisms during male germ-cell development in mice. *PLoS genetics* 13, e1006926. [PubMed: 28749988]
- Jegou B (1993). The Sertoli-germ cell communication network in mammals. *Int Rev Cytol* 147, 25–96. [PubMed: 8225836]
- Johnston DS, Wright WW, Dicaneloro P, Wilson E, Kopf GS, and Jelinisky SA (2008). Stage-specific gene expression is a fundamental characteristic of rat spermatogenic cells and Sertoli cells. *Proceedings of the National Academy of Sciences of the United States of America* 105, 8315–8320. [PubMed: 18544648]
- Kissel H, Timokhina I, Hardy MP, Rothschild G, Tajima Y, Soares V, Angeles M, Whitlow SR, Manova K, and Besmer P (2000). Point mutation in kit receptor tyrosine kinase reveals essential roles for kit signaling in spermatogenesis and oogenesis without affecting other kit responses. *The EMBO journal* 19, 1312–1326. [PubMed: 10716931]
- Kliesch S, Penttila TL, Gromoll J, Saunders PT, Nieschlag E, and Parvinen M (1992). FSH receptor mRNA is expressed stage-dependently during rat spermatogenesis. *Molecular and cellular endocrinology* 84, R45–49. [PubMed: 1587387]
- Komai Y, Tanaka T, Tokuyama Y, Yanai H, Ohe S, Omachi T, Atsumi N, Yoshida N, Kumano K, Hisha H, et al. (2014). Bmi1 expression in long-term germ stem cells. *Scientific reports* 4, 6175. [PubMed: 25146451]
- Kubota H, Avarbock MR, and Brinster RL (2004). Culture conditions and single growth factors affect fate determination of mouse spermatogonial stem cells. *Biol Reprod* 71, 722–731. [PubMed: 15115718]
- Lesch BJ, Dokshin GA, Young RA, McCarrey JR, and Page DC (2013). A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. *Proceedings of the National Academy of Sciences of the United States of America* 110, 16061–16066. [PubMed: 24043772]
- Luoh SW, Bain PA, Polakiewicz RD, Goodheart ML, Gardner H, Jaenisch R, and Page DC (1997). Zfx mutation results in small animal size and reduced germ cell number in male and female mice. *Development (Cambridge, England)* 124, 2275–2284.
- Machanic P, and Bailey TL (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics (Oxford, England)* 27, 1696–1697.
- Macosko EZ, Basu A, Satija R, Nemes J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161, 1202–1214. [PubMed: 26000488]
- Maekawa M, Kamimura K, and Nagano T (1996). Peritubular myoid cells in the testis: their structure and function. *Arch Histol Cytol* 59, 1–13. [PubMed: 8727359]
- Maresz K, Ponomarev ED, Barteneva N, Tan Y, Mann MK, and Dittel BN (2008). IL-13 induces the expression of the alternative activation marker Ym1 in a subset of testicular macrophages. *Journal of reproductive immunology* 78, 140–148. [PubMed: 18329106]
- Matson CK, Murphy MW, Griswold MD, Yoshida S, Bardwell VJ, and Zarkower D (2010). The mammalian doublesex homolog DMRT1 is a transcriptional gatekeeper that controls the mitosis versus meiosis decision in male germ cells. *Developmental cell* 19, 612–624. [PubMed: 20951351]
- Matzuk MM, and Lamb DJ (2008). The biology of infertility: research advances and clinical challenges. *Nat Med* 14, 1197–1213. [PubMed: 18989307]

- McKee BD, and Handel MA (1993). Sex chromosomes, recombination, and chromatin conformation. *Chromosoma* 102, 71–80. [PubMed: 8432196]
- McKusick-Nathans Institute of Genetic Medicine, J.H.U.B. MD) (2017). Online Mendelian Inheritance in Man, OMIM®.
- Meistrich ML, and Hess RA (2013). Assessment of spermatogenesis through staging of seminiferous tubules. *Methods in molecular biology* (Clifton, NJ) 927, 299–307.
- Meng X, Lindahl M, Hyvonen ME, Parvinen M, de Rooij DG, Hess MW, Raatikainen-Ahokas A, Sainio K, Rauvala H, Lakso M, et al. (2000). Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. *Science* (New York, NY) 287, 1489–1493.
- Moore A, and Morris ID (1993). The involvement of insulin-like growth factor-I in local control of steroidogenesis and DNA synthesis of Leydig and non-Leydig cells in the rat testicular interstitium. *J Endocrinol* 138, 107–114. [PubMed: 7531750]
- Morales C, and Griswold MD (1987). Retinol-induced stage synchronization in seminiferous tubules of the rat. *Endocrinology* 121, 432–434. [PubMed: 3595524]
- Mueller JL, Mahadevaiah SK, Park PJ, Warburton PE, Page DC, and Turner JM (2008). The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nature genetics* 40, 794–799. [PubMed: 18454149]
- Nakagawa T, Sharma M, Nabeshima Y, Braun RE, and Yoshida S (2010). Functional hierarchy and reversibility within the murine spermatogenic stem cell compartment. *Science* (New York, NY) 328, 62–67.
- Nalbandian A, Dettin L, Dym M, and Ravindranath N (2003). Expression of vascular endothelial growth factor receptors during male germ cell differentiation in the mouse. *Biol Reprod* 69, 985–994. [PubMed: 12773425]
- Neill DR, Wong SH, Bellosi A, Flynn RJ, Daly M, Langford TK, Bucks C, Kane CM, Fallon PG, Pannell R, et al. (2010). Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. *Nature* 464, 1367–1370. [PubMed: 20200518]
- Nishimune Y, Aizawa S, and Komatsu T (1978). Testicular germ cell differentiation in vivo. *Fertility and sterility* 29, 95–102. [PubMed: 23321]
- O'Shaughnessy PJ, Hu L, and Baker PJ (2008). Effect of germ cell depletion on levels of specific mRNA transcripts in mouse Sertoli cells and Leydig cells. *Reproduction* 135, 839–850. [PubMed: 18390686]
- Oatley JM, Avarbock MR, Telaranta AI, Fearon DT, and Brinster RL (2006). Identifying genes important for spermatogonial stem cell self-renewal and survival. *Proceedings of the National Academy of Sciences of the United States of America* 103, 9524–9529. [PubMed: 16740658]
- Oatley JM, Oatley MJ, Avarbock MR, Tobias JW, and Brinster RL (2009). Colony stimulating factor 1 is an extrinsic stimulator of mouse spermatogonial stem cell self-renewal. *Development* 136, 1191–1199. [PubMed: 19270176]
- Ohbo K, Yoshida S, Ohmura M, Ohneda O, Ogawa T, Tsuchiya H, Kuwana T, Kehler J, Abe K, Scholer HR, et al. (2003). Identification and characterization of stem cells in prepubertal spermatogenesis in mice. *Developmental biology* 258, 209–225. [PubMed: 12781694]
- Parvinen M, Pelto-Huikko M, Soder O, Schultz R, Kaipia A, Mali P, Toppari J, Hakovirta H, Lonnerberg P, Ritzen EM, et al. (1992). Expression of beta-nerve growth factor and its receptor in rat seminiferous epithelium: specific function at the onset of meiosis. *The Journal of cell biology* 117, 629–641. [PubMed: 1315318]
- Pech MF, Garbuzov A, Hasegawa K, Sukhwani M, Zhang RJ, Benayoun BA, Brockman SA, Lin S, Brunet A, Orwig KE, et al. (2015). High telomerase is a hallmark of undifferentiated spermatogonia and is required for maintenance of male germline stem cells. *Genes & development* 29, 2420–2434. [PubMed: 26584619]
- Pepling ME, de Cuevas M, and Spradling AC (1999). Germline cysts: a conserved phase of germ cell development? *Trends Cell Biol* 9, 257–262. [PubMed: 10370240]
- Phillips BT, Gassei K, and Orwig KE (2010). Spermatogonial stem cell regulation and spermatogenesis. *Philos Trans R Soc Lond B Biol Sci* 365, 1663–1678. [PubMed: 20403877]
- Piras V, Tomita M, and Selvarajoo K (2014). Transcriptome-wide variability in single embryonic development cells. *Scientific reports* 4, 7137. [PubMed: 25409746]

- Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, and Trapnell C (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nature methods* 14, 979–982. [PubMed: 28825705]
- Raj A, and Tyagi S (2010). Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. *Methods in enzymology* 472, 365–386. [PubMed: 20580972]
- Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, and Tyagi S (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nature methods* 5, 877–879. [PubMed: 18806792]
- Rantanen A, and Larsson NG (2000). Regulation of mitochondrial DNA copy number during spermatogenesis. *Human reproduction (Oxford, England)* 15 Suppl 2, 86–91.
- Rodriguez A, and Laio A (2014). Machine learning. Clustering by fast search and find of density peaks. *Science (New York, NY)* 344, 1492–1496.
- Satija R, Farrell JA, Gennert D, Schier AF, and Regev A (2015). Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 33, 495–502. [PubMed: 25867923]
- Schrans-Stassen BH, van de Kant HJ, de Rooij DG, and van Pelt AM (1999). Differential expression of c-kit in mouse undifferentiated and differentiating type A spermatogonia. *Endocrinology* 140, 5894–5900. [PubMed: 10579355]
- Sharpe RM (1986). Paracrine control of the testis. *Clin Endocrinol Metab* 15, 185–207. [PubMed: 3514003]
- Shekhar K, Lapan SW, Whitney IE, Tran NM, Macosko EZ, Kowalczyk M, Adiconis X, Levin JZ, Nemesh J, Goldman M, et al. (2016). Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell* 166, 1308–1323 e1330. [PubMed: 27565351]
- Shin J, Berg DA, Zhu Y, Shin JY, Song J, Bonaguidi MA, Enikolopov G, Nauen DW, Christian KM, Ming GL, et al. (2015). Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. *Cell Stem Cell* 17, 360–372. [PubMed: 26299571]
- Shinohara T, Orwig KE, Avarbock MR, and Brinster RL (2000). Spermatogonial stem cell enrichment by multiparameter selection of mouse testis cells. *Proceedings of the National Academy of Sciences of the United States of America* 97, 8346–8351. [PubMed: 10900001]
- Sleutels F, Soochit W, Bartkuhn M, Heath H, Dienstbach S, Bergmaier P, Franke V, Rosa-Garrido M, van de Nobelen S, Caesar L, et al. (2012). The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner. *Epigenetics & chromatin* 5, 8. [PubMed: 22709888]
- Smith LB, and Walker WH (2014). The regulation of spermatogenesis by androgens. *Semin Cell Dev Biol* 30, 2–13. [PubMed: 24598768]
- Snyder EM, Davis JC, Zhou Q, Evanoff R, and Griswold MD (2011). Exposure to retinoic acid in the neonatal but not adult mouse results in synchronous spermatogenesis. *Biol Reprod* 84, 886–893. [PubMed: 21228214]
- Soderstrom KO, and Parvinen M (1976). RNA synthesis in different stages of rat seminiferous epithelial cycle. *Molecular and cellular endocrinology* 5, 181–199. [PubMed: 955250]
- Solari AJ (1974). The behavior of the XY pair in mammals. *International review of cytology* 38, 273–317. [PubMed: 4854664]
- Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, Barthes P, Kokkinaki M, Nef S, Gnirke A, et al. (2013). Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep* 3, 2179–2190. [PubMed: 23791531]
- Spits H, and Di Santo JP (2011). The expanding family of innate lymphoid cells: regulators and effectors of immunity and tissue remodeling. *Nat Immunol* 12, 21–27. [PubMed: 21113163]
- Stephenson W, Donlin LT, Butler A, Rozo C, Bracken B, Rashidfarrokhi A, Goodman SM, Ivashkiv LB, Bykerk VP, Orange DE, et al. (2018). Single-cell RNA-seq of rheumatoid arthritis synovial tissue using low-cost microfluidic instrumentation. *Nat Commun* 9, 791. [PubMed: 29476078]
- Stevant I, Neirijnck Y, Borel C, Escoffier J, Smith LB, Antonarakis SE, Dermitzakis ET, and Nef S (2018). Deciphering Cell Lineage Specification during Male Sex Determination with Single-Cell RNA Sequencing. *Cell Rep* 22, 1589–1599. [PubMed: 29425512]
- Sun F, Xu Q, Zhao D, and Degui Chen C (2015). Id4 Marks Spermatogonial Stem Cells in the Mouse Testis. *Scientific reports* 5, 17594. [PubMed: 26621350]

- Suzuki H, Ahn HW, Chu T, Bowden W, Gassei K, Orwig K, and Rajkovic A (2012). SOHLH1 and SOHLH2 coordinate spermatogonial differentiation. *Developmental biology* 361, 301–312. [PubMed: 22056784]
- Suzuki T, Kosaka-Suzuki N, Pack S, Shin DM, Yoon J, Abdullaev Z, Pugacheva E, Morse HC 3rd, Loukinov D, and Lobanenko V (2010). Expression of a testis-specific form of Gal3st1 (CST), a gene essential for spermatogenesis, is regulated by the CTCF paralogous gene BORIS. *Molecular and cellular biology* 30, 2473–2484. [PubMed: 20231363]
- Tanay A, and Regev A (2017). Scaling single-cell genomics from phenomenology to mechanism. *Nature* 541, 331–338. [PubMed: 28102262]
- Teschendorff AE, and Enver T (2017). Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. *Nature communications* 8, 15599.
- Tourtellotte WG, Nagarajan R, Bartke A, and Milbrandt J (2000). Functional compensation by Egr4 in Egr1-dependent luteinizing hormone regulation and Leydig cell steroidogenesis. *Molecular and cellular biology* 20, 5261–5268. [PubMed: 10866682]
- Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, and Rinn JL (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 32, 381–386. [PubMed: 24658644]
- Ueno H, and Mori H (1990). Morphometrical analysis of Sertoli cell ultrastructure during the seminiferous epithelial cycle in rats. *Biology of reproduction* 43, 769–776. [PubMed: 2291913]
- Van Beek ME, and Meistrich ML (1990). A method for quantifying synchrony in testes of rats treated with vitamin A deprivation and readministration. *Biol Reprod* 42, 424–431. [PubMed: 2340329]
- van der Maaten LJP, and Hinton GE (2008). Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research* 9, 2579–2605.
- Wang J, Tang C, Wang Q, Su J, Ni T, Yang W, Wang Y, Chen W, Liu X, Wang S, et al. (2017). NRF1 coordinates with DNA methylation to regulate spermatogenesis. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 31, 4959–4970. [PubMed: 28754714]
- Wright WW, Smith L, Kerr C, and Charron M (2003). Mice that express enzymatically inactive cathepsin L exhibit abnormal spermatogenesis. *Biology of reproduction* 68, 680–687. [PubMed: 12533435]
- Xu J, Lamouille S, and Derynck R (2009). TGF-beta-induced epithelial to mesenchymal transition. *Cell Res* 19, 156–172. [PubMed: 19153598]
- Yan J (2016). som: Self-Organizing Map.
- Yang D, Yang W, Tian Z, van Velkinburgh JC, Song J, Wu Y, and Ni B (2016). Innate lymphoid cells as novel regulators of obesity and its-associated metabolic dysfunction. *Obesity reviews : an official journal of the International Association for the Study of Obesity* 17, 485–498. [PubMed: 26948388]
- Yoshida S, Sueno M, and Nabeshima Y (2007). A vasculature-associated niche for undifferentiated spermatogonia in the mouse testis. *Science (New York, NY)* 317, 1722–1726.
- Zhang H, Akman HO, Smith EL, Zhao J, Murphy-Ullrich JE, and Batuman OA (2003). Cellular response to hypoxia involves signaling via Smad proteins. *Blood* 101, 2253–2260. [PubMed: 12411310]
- Zhang T, Murphy MW, Gearhart MD, Bardwell VJ, and Zarkower D (2014). The mammalian Doublesex homolog DMRT6 coordinates the transition between mitotic and meiotic developmental programs during spermatogenesis. *Development (Cambridge, England)* 141, 3662–3671.
- Zhang T, and Zarkower D (2017). DMRT proteins and coordination of mammalian spermatogenesis. *Stem cell research* 24, 195–202. [PubMed: 28774758]
- Zhou Q, Nie R, Li Y, Friel P, Mitchell D, Hess RA, Small C, and Griswold MD (2008). Expression of stimulated by retinoic acid gene 8 (Stra8) in spermatogenic cells induced by retinoic acid: an in vivo study in vitamin A-sufficient postnatal murine testes. *Biology of reproduction* 79, 35–42. [PubMed: 18322276]

Zimmermann C, Stevant I, Borel C, Conne B, Pitetti JL, Calvel P, Kaessmann H, Jegou B, Chalmel F, and Nef S (2015). Research resource: the dynamic transcriptional profile of sertoli cells during the progression of spermatogenesis. *Molecular endocrinology* (Baltimore, Md) 29, 627–642.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Highlights

1. Analysis of ~35K cells identifies known and unexpected mouse testicular cell-types.
2. Germ cell development includes discrete states followed by a continuous trajectory.
3. Differential gene expression identifies novel regulators of spermatogenesis.
4. Four spermatogonia subtypes and nine Sertoli cell subtypes map to known stages.

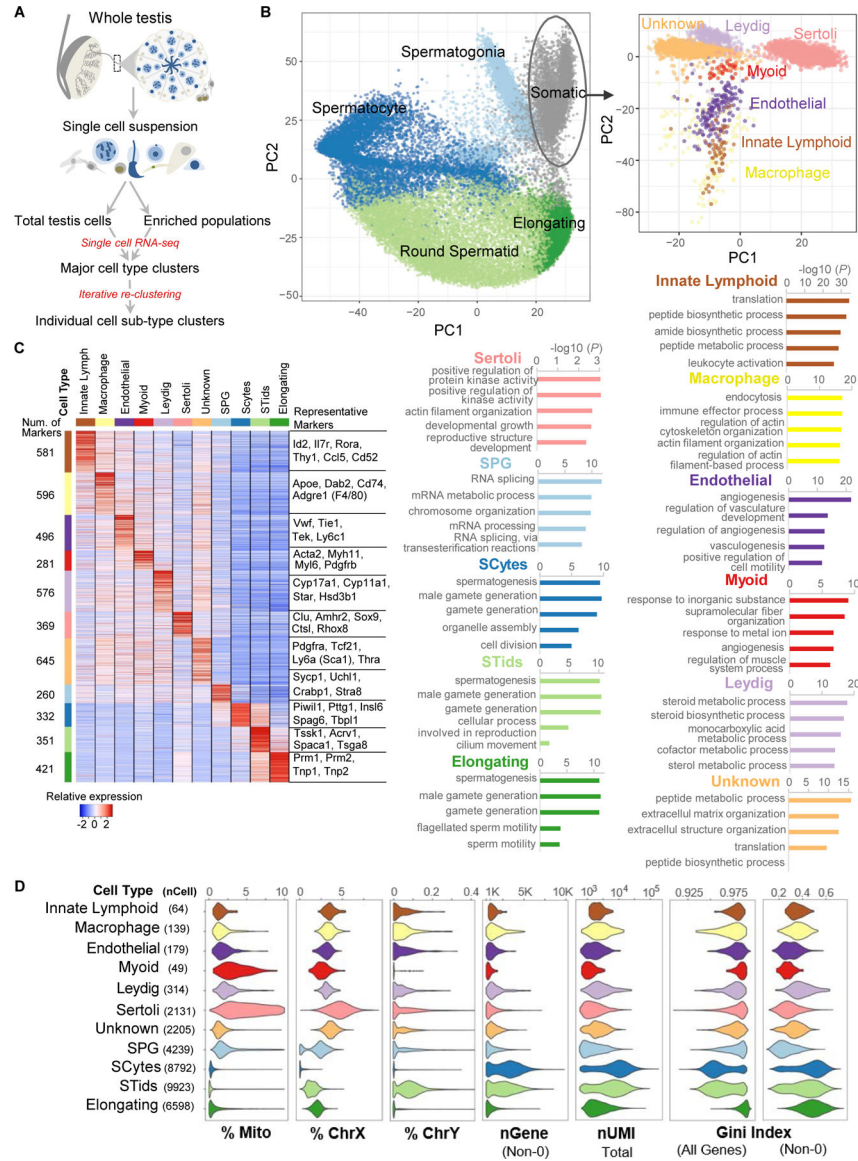


Figure 1. Overview of major cell types and cellular attributes inferred from single-cell RNA-seq analyses of the mouse testis. (A) Schematic overview of data collection and iterative clustering approach. (B) Cellular heterogeneity at the highest levels. Left: principal component analysis of all ~35K cells post-QC reveals 5 major clusters, corresponding to four germ cell and one somatic cell cluster. Right: focused re-clustering of the 5,081 somatic cells identifies seven cell types: macrophages, endothelial, myoid, Leydig, Sertoli, innate lymphoid type II cells, and a previously unexpected mesenchymal cell type (“unknown”). (C) Marker genes and their top 5 gene ontologies, highlighting salient biological functions of the major cell types. Note – in the heatmap each marker gene is standardized over the 11 cluster centroids and ordered by cell type. (D) Distribution profiles of per-cell attributes compared across the 11 cell types. From left to right: %Mito, percent of mitochondria transcripts in the overall transcriptome; %ChrX and %ChrY, percent of X and Y chromosome transcripts, respectively; nGene, total number of detected genes in a given cell; nUMI, total number of Unique Molecular

Identifiers (UMI) in a given cell, a.k.a the “cell size factor”; Gini Index, a measure of gene expression inequality in each cell using either all ~35K genes expressed in at least one cell (left) or only the detected genes (with non-zero counts) for that cell (right).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

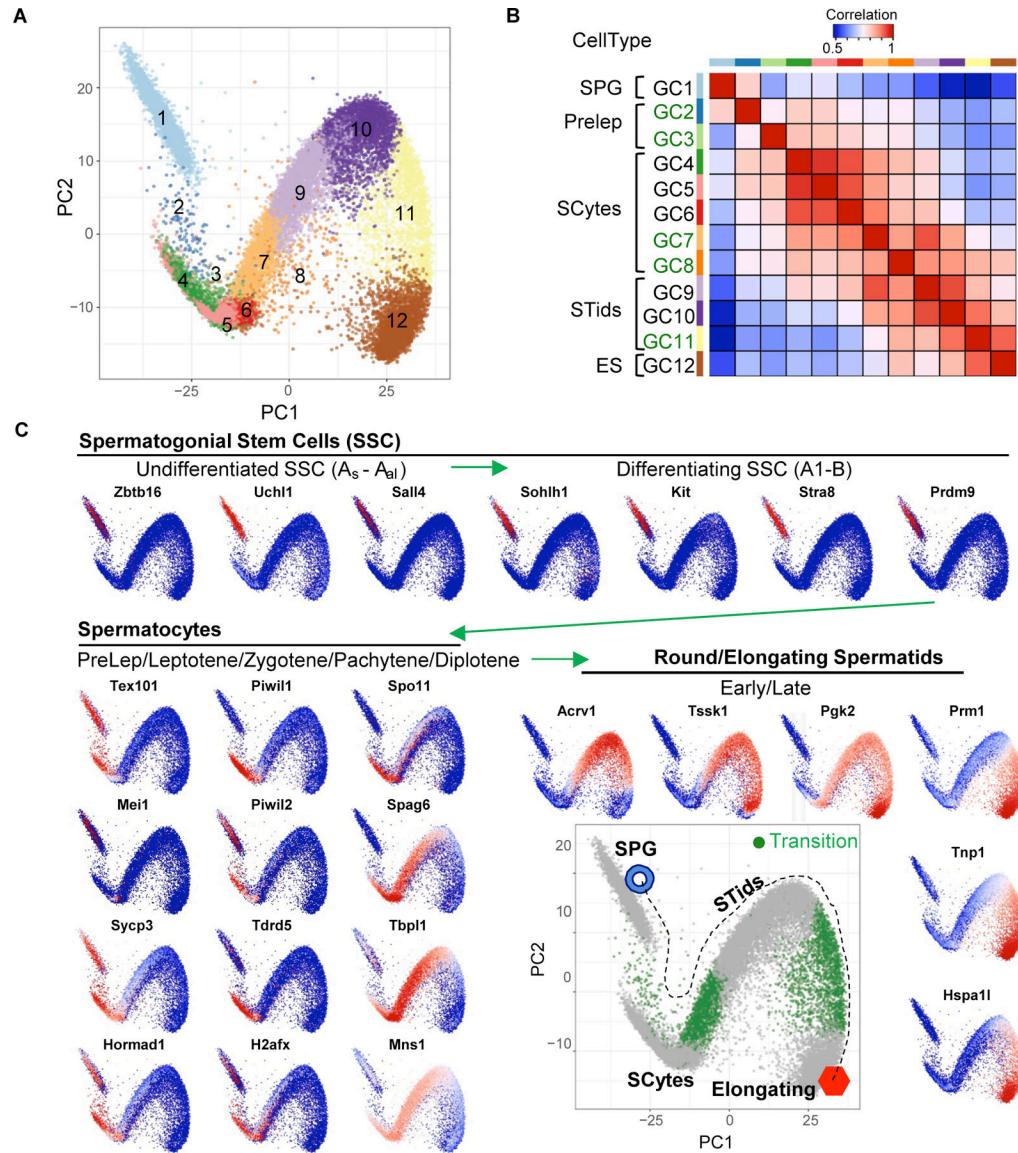


Figure 2. Adult germ cell development exhibits both discrete states and continuous developmental transitions.

(A) Principal component plot of 20,646 germ cells with >1,000 detected genes, colored by assignment to 12 clusters determined by unbiased clustering. (B) Pairwise rank correlation matrix among the 12 cluster centroids, showing that Clusters GC1–3 are relatively isolated whereas the other 9 GC clusters form a gradual series of transitions. (C) Biological annotation of the 12 germ cell clusters using genes of known, stage-specific expression. The seven markers in the top row suggest that cells in GC1 correspond to spermatogonia (SPG) – comprised of undifferentiated and differentiating spermatogonia (see Figure 4 for zoomed in clustering of spermatogonia). GC2–3 likely contain rare cells transitioning into meiosis. According to the 12 markers shown in the lower left panel, GC4–8 correspond to spermatocyte (SCytes). Whereas genes in the right panel suggest that cells in GC9–12 correspond to round spermatids (STids, Clusters GC9–11) and elongated spermatids (ES, Cluster GC12). Major biological transitions are highlighted in green.

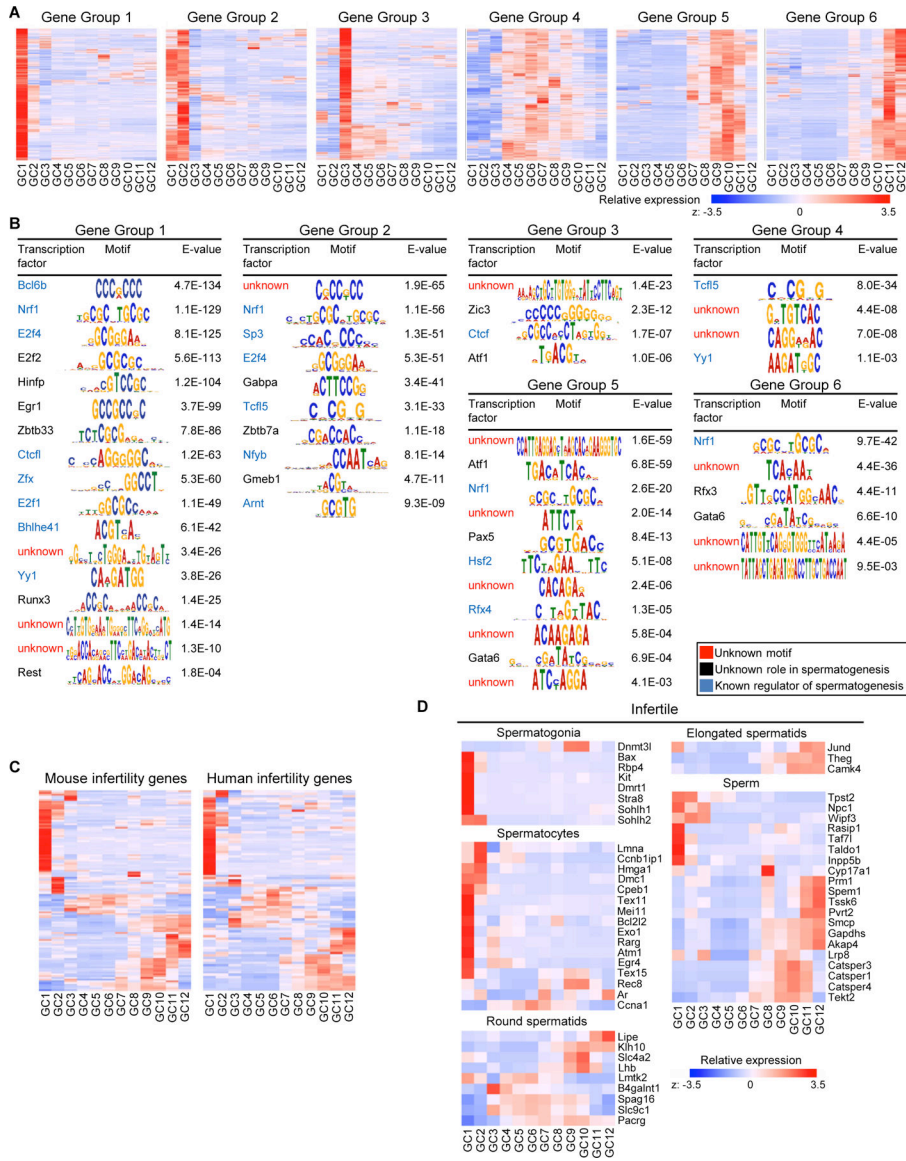


Figure 3. Gene expression dynamics along the germ cell differentiation trajectory. (A) Unsupervised K-means clustering (k=6) of 8,583 highly variable genes across the 12 germ cell cluster centroids yields six groups of genes with distinct expression patterns. From left to right are six heatmaps of scaled expression levels across the 12 centroids, showing wave-like progression of gene expression from Group 1 genes, which are highly expressed in spermatogonia (germ cell cluster GC1), to Group 6 genes, highly expressed in elongated spermatids (germ cell cluster GC12). (B) Transcription factor motifs significantly enriched (E-value < 0.01) within +/-1kb of the transcriptional start site of the six groups of genes. (C) Gene expression heatmaps of 187 mouse male-infertility genes (left) and 234 human infertility genes (right) over the 12 germ cell clusters, highlighting a significant proportion of mammalian infertility genes have peak expression in spermatogonia. (D) Gene expression heatmaps of mouse infertility genes grouped by the five known stages of germ cell arrest,

showing that genes causing arrest in a particular stage tend to be expressed at high levels in the same or an earlier stage.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

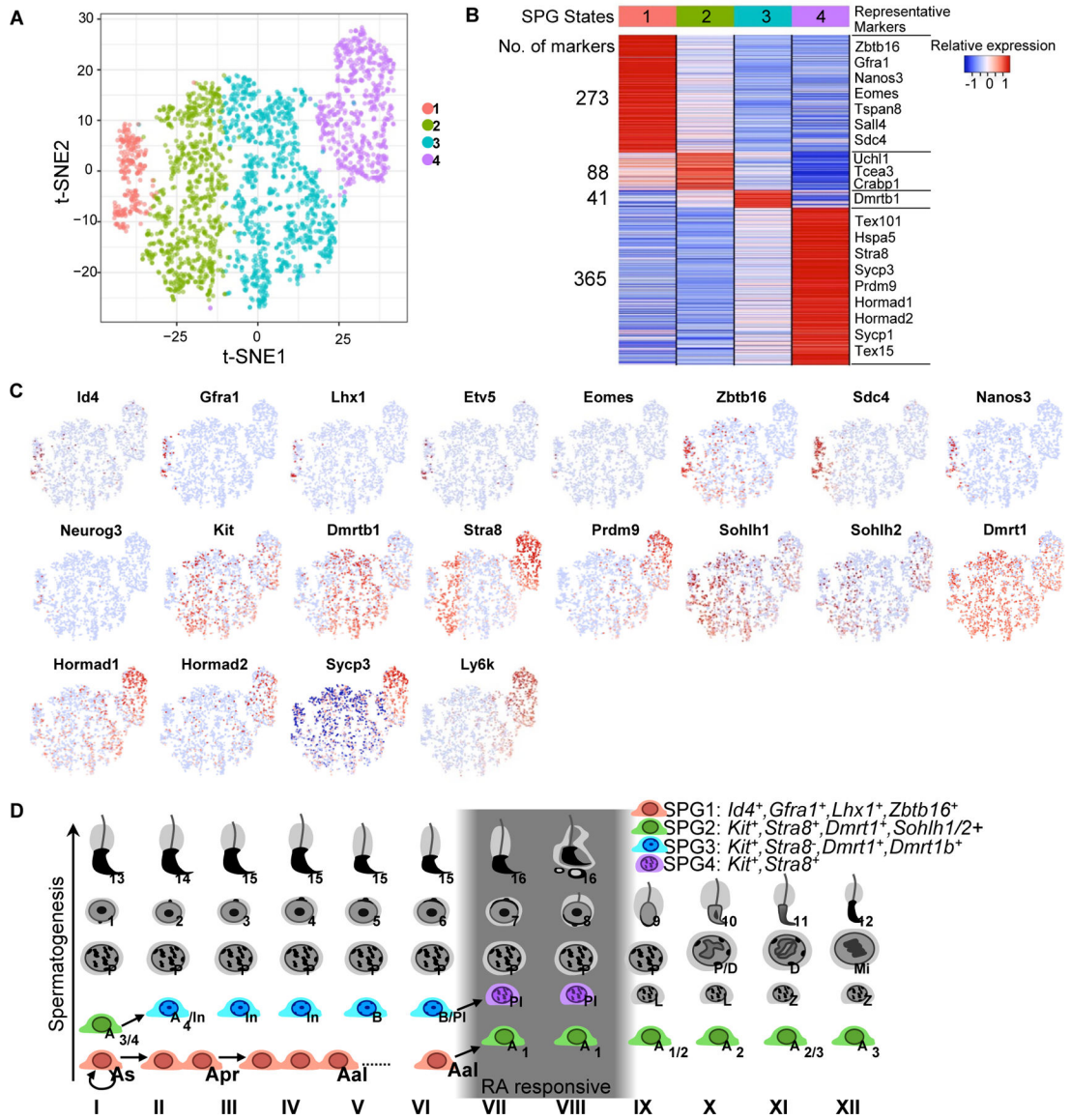


Figure 4. Heterogeneity among spermatogonia cells supports 4 recognized subtypes: SPG1-SPG4.
 (A) Focused re-clustering of 2,484 spermatogonia cells with >1,000 UMIs reveals 4 biological subtypes, as shown in the t-SNE plot. (B) Heatmap of differentially expressed marker genes, obtained by comparing each subtype against the other three ($p < 0.01$; fold change > 1.5). (C) Per-cell expression level of known or novel markers of the four spermatogonia states visualized in t-SNE space. (D) Summary schematic depicting the position of spermatogonia subtypes across stages of the mouse seminiferous epithelial cycle. Illustration is modified from (Ahmed EA and de Rooij DG, 2009; Meistrich ML1 and Hess RA, 2013).

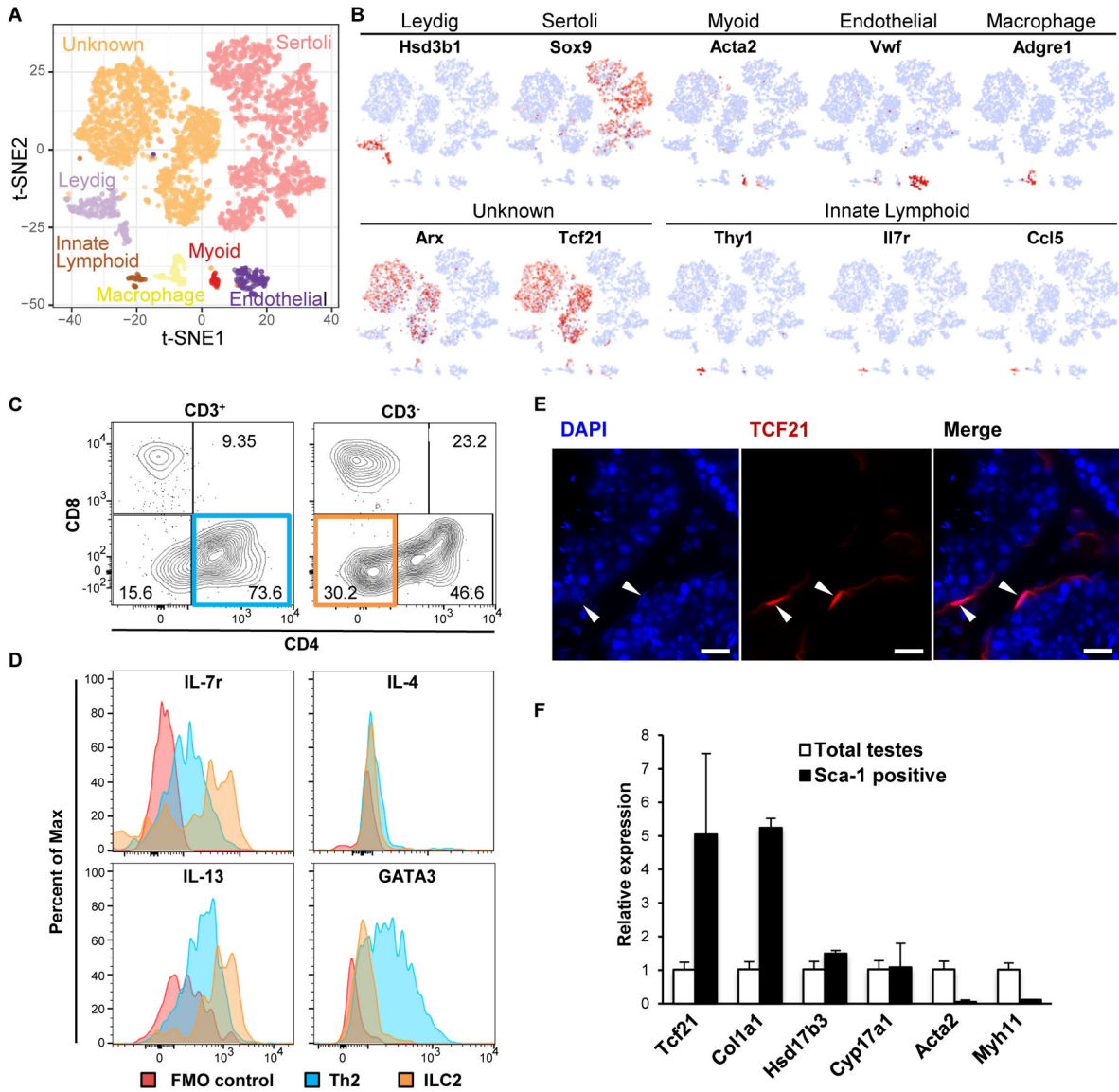


Figure 5. Identification of known and new somatic cell types in the testis. (A) Focused re-clustering of 5,081 somatic cells revealed seven distinct cell types as shown in t-SNE space. Note that the relative cell number proportions illustrated in TSNE plots is not representative of *in vivo* proportions, since many of the somatic populations required genetic or molecular enrichment experiments prior to Drop-seq analysis. (B) Cell-type specific expression of selected maker genes shown in t-SNE space. (C) Identification of resident ILCII population in the testis using flow cytometry. T_H2 are designated as $CD3^+/CD4^+/CD8^-$ and ILCII cells are $CD3^-/CD8^-/CD4^-$. (D) Further validation of the ILCII population can be achieved using known cell surface or intracellular markers (IL7R, GATA3, IL-13, and IL-4). (E) Localization of the $Tcf21^+$ mesenchymal cell population in the testis by genetic labeling using $Tcf21$ -creERT2; tdTomato mice. White arrowheads mark $Tcf21^+$ cells surrounding seminiferous tubules. (F) Validation of $Tcf21$ and $Col1a1$ mRNA expression in $Sca1^+$ cells by real time qRT-PCR. The $Sca1^+$ cells are depleted of Leydig

cells markers (*Hsd3b1* and *Cyp17a1*), and myoid cell markers (*Myh11* and *Acta2*). Data represent average \pm SD.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

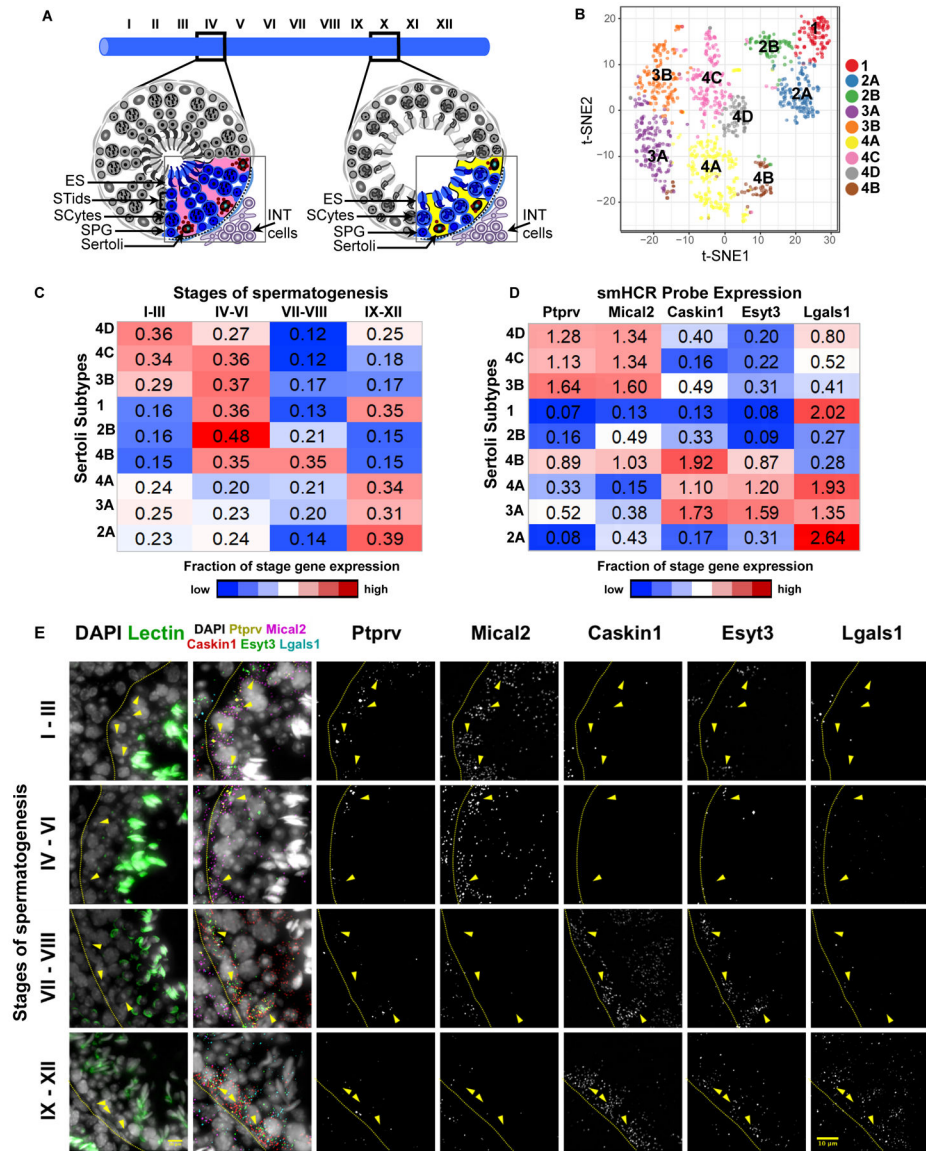


Figure 6. Functional subtypes of Sertoli cells map to spatially defined seminiferous tubule stages.

(A) Schematic illustrating Sertoli cell heterogeneity across the 12 stages of the mouse seminiferous epithelial cycle. (B) Unbiased clustering of Sertoli cells reveals four major functional types (SER-1–4), which can be further divided into nine subtypes (named with a letter suffix, e.g., SER-2A/B for the two subtypes obtained from SER-2). (C) Comparison of the nine transcriptome-based Sertoli subtypes with four stage-specific Sertoli cell enriched marker gene lists identified by microarrays from tubule segments (Hasegawa and Saga, 2012; Wright et al., 2003). Specifically, we calculated the relative fraction of Stages I-III, IV-VI, VII-VIII, or IX-XII genes across the 9 Sertoli cell subtypes. This fraction is calculated for every cell, then averaged in each of the nine molecular clusters, forming the 9-by-4 matrix. (D) Heatmap of expression levels for the five probes designed for smHCR across 9 Sertoli subtype centroids. The values displayed are natural log-transformed cluster centroid average expression values for each gene. The marker probes chosen for smHCR

enrich in multiple Sertoli cell subtypes and aim to examine whether the Sertoli cell subtypes derived from a major cluster do or do not colocalize in situ. (E) smHCR reveals stage-specific expression of five Sertoli cell marker genes. For each row of imaging panels, left panel shows seminiferous tubule staging determined by the pattern of acrosome staining with Lectin PNA; second to left panel shows the combined RNA transcripts by smHCR; right five panels show the isolated signal from each probe. Arrowheads indicate Sertoli cell nuclei. Dashed lines represent tubule borders.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

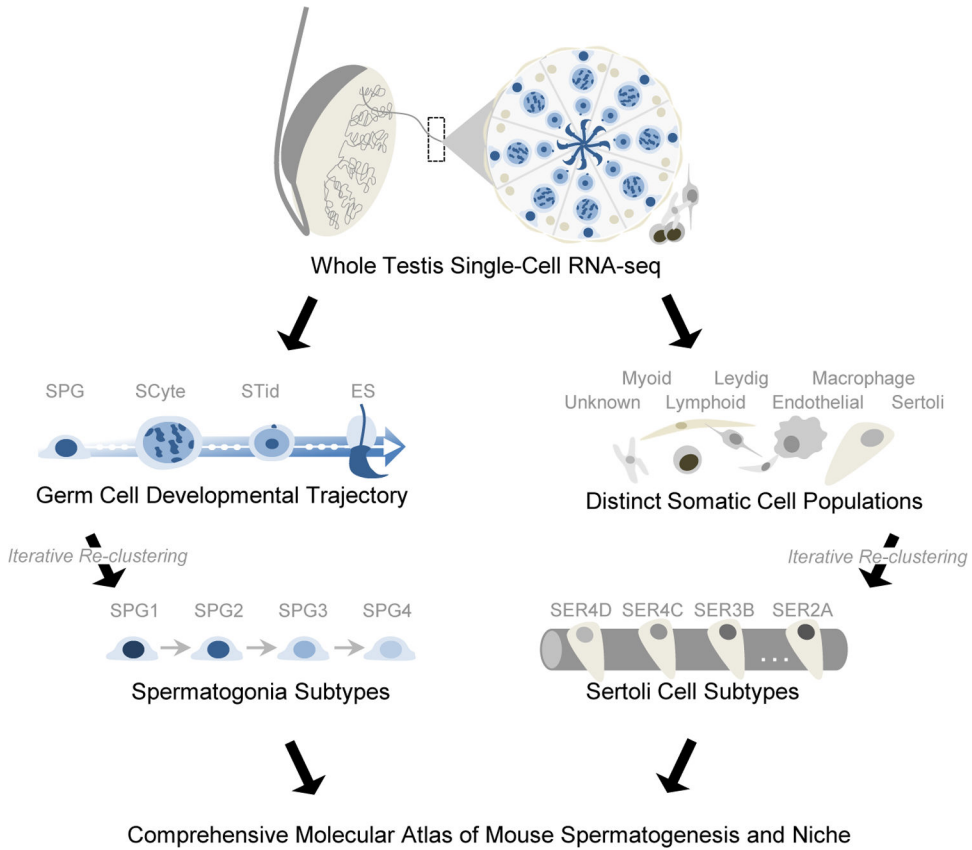


Figure 7. Overview of the comprehensive cellular atlas of mouse spermatogenesis and testis niche.

Summary schematic of the major findings from the analysis of >35K single-cell RNA-seq profiles. On the Left, our study demonstrates for the first time the full developmental trajectory of germ cell development from spermatogonia to elongated spermatids. The transition from spermatogonia to spermatocytes involves discrete developmental transitions, whereas, the progression from spermatocytes to elongating spermatids is continuous with no stable intermediate states. Focused re-clustering of spermatogonia further define transitions between undifferentiated and differentiated stem cells. On the Right, we identify all major somatic cell types within the testis, as well as two previously uncharacterized populations (innate lymphoid type 2 cells and an unknown mesenchymal cell type). Focused re-clustering of Sertoli cells uncovers significant heterogeneity which can be linked biologically to cycling stages of the seminiferous epithelium. Taken together, these findings represent a powerful new resource to the community for studying the cellular and molecular heterogeneity of the testis and spermatogenesis program in unprecedented resolution.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rat monoclonal anti-CD90 [30-H12] (FITC)	Abcam	Cat#ab62009; RRID: AB_940927
Mouse monoclonal anti-CD90.1 (HIS51), PerCP-Cyanine5.5	ThermoFisher Scientific	Cat#45090082; RRID: AB_2573662
Rat monoclonal anti-Ly6A/E (Sca-1) (D7), PerCP-Cyanine5.5	ThermoFisher Scientific	Cat#45-5981-82; RRID: AB_914372
Streptavidin, Alexa Fluor™ 488 conjugate	ThermoFisher Scientific	Cat#S11223; RRID: AB_2336881
PE/Cy7 anti-mouse CD3 Antibody	Biolegend	Cat#100219; RRID:AB_1732068
Brilliant Violet 510™ anti-mouse CD4 Antibody	Biolegend	Cat#100449; RRID:AB_2564587
Brilliant Violet 570™ anti-mouse CD8a Antibody	Biolegend	Cat#100739; RRID:AB_524958
PE/Cy5 anti-mouse CD127 (IL-7R α) Antibody	Biolegend	Cat#135015; RRID:AB_1937262
PE anti-mouse IL-4 Antibody	Biolegend	Cat#504103; RRID:AB_315317
APC anti-mouse IL-13 Antibody	Novus	Cat#011818
Brilliant Violet 421™ anti-GATA3 Antibody	Biolegend	Cat#653813; RRID:AB_2563220
APC anti-mouse CD45 Antibody	BD Bioscience	Cat#561018; RRID:AB_10584326
Chemicals, Peptides, and Recombinant Proteins		
Deoxyribonuclease I	Worthington Biochemical Corp.	Cat#LS002139
Collagenase Type IA	Sigma	Cat#C9891
Advanced DMEM:F12 media	ThermoFisher Scientific	Cat#12634010
Trypsin	ThermoFisher Scientific	Cat#27250018
autoMACS Rinsing Solution	Miltenyi Biotec	Cat#130-091-222
MACS BSA Stock Solution	Miltenyi Biotec	Cat#130-091-376
Hoechst 33342	ThermoFisher Scientific	Cat#H3570
Propidium Iodide	ThermoFisher Scientific	Cat#P3566
DAPI	Sigma	Cat#D9542
FBS	ThermoFisher Scientific	Ca#10437010
Tamoxifen	Sigma	Cat#T5648
Collagenase D	Sigma	Cat#11088858001
Sucrose, Rnase and Dnase free	Amresco	Cat#0335
p-Phenylenediamine	Sigma	Cat#p6001
Lectin PNA, Alexa Fluor® 647 Conjugate	ThermoFisher Scientific	Cat#L32460
Paraformaldehyde	EMD Millipore	Cat#818715
Bovine Serum Albumin, Fraction V	VWR	Cat#97061-422
ProLong™ Gold Antifade Mountant	ThermoFisher Scientific	Cat#P36930
Formamide, Deionized	Ambion	Cat#AM9342
tRNA from Baker's Yeast	Sigma	Cat#10109495001
Ribonucleoside Vanadyl Complex	New England BioLabs	Cat#S1402S
UltraPure™ BSA	ThermoFisher Scientific	Cat#AM2618

Ficoll PM-400	Sigma	Cat#F5415
Sarkosyl, Sodium Salt Solution	Sigma	Cat#L7414
1H,1H,2H,2H-Perfluoro-1-octanol	Sigma	Cat#370533
dNTP Mix	ThermoFisher Scientific	Cat#R0192
NxGen RNase Inhibitor	Lucigen	Cat#30-281-1
Maxima H Minus Reverse Transcriptase	ThermoFisher Scientific	Cat#EP0753
Exonuclease I	New England BioLabs	Cat#M0293S
SuperScript™ III First-Strand Synthesis System	ThermoFisher Scientific	Cat#18080-051
Power SYBR™ Green PCR Master Mix	ThermoFisher Scientific	Cat#4367659
GolgiStop™	BD Bioscience	Cat#554724
Agencourt AMPure XP Beads	Beckman Coulter	Cat#A63880
Tween-20	ThermoFisher Scientific	Cat#00-3005
Heparin	Sigma	Cat#H3393
Denhard s solution	ThermoFisher Scientific	Cat#750018
Citric Acid	Sigma	Cat#791725
Triton X-100	Sigma	Cat#T8787
Dextran Sulfate	Sigma	Cat#D6001
Salmon Sperm DNA	ThermoFisher Scientific	Cat#15632-011
SSC Buffer	ThermoFisher Scientific	Cat#15557-044
Ionomycin	Cell Signaling Technology	Cat#9995
PMA	Sigma	Cat#P1585
Fluorescent Nanodiamonds	Adamas Nano	Cat#NDNV100nmHi10ml
Critical Commercial Assays		
Rat monoclonal anti-CD117 microbeads	Miltenyl Biotec	Cat#130-091-224
Rat monoclonal anti-CD90.2 microbeads	Miltenyl Biotec	Cat#130-049-101
QuadroMACS Starting Kit (LS)	Miltenyl Biotec	Cat#130-091-051
KAPA HiFi HotStart ReadyMix PCR Kit	Kapa Biosystems	Cat#KK2602
Nextera XT DNA SMP Prep Kit	Illumina	Cat#FC-131-1096
Deposited Data		
Raw data files for RNA-sequencing	This paper or NCBI Gene Expression Omnibus	GEO: GSE112393
Experimental Models: Organisms/Strains		
Mouse: C57BL/6J	The Jackson Laboratory	JAX: 000664
Mouse: Gfra1-CreERT2	Hara et al., 2014	N/A
Mouse: B6.129(Cg)-Gt(ROSA)26Sortm4(ACTB-tdTomato,-EGFP)Luo/J	The Jackson Laboratory	JAX: 007676
Mouse: 129S.FVB-Tg(Amh-cre)8815Reb/J	The Jackson Laboratory	JAX: 007915
Mouse: Tg(Sox9-EGFP)EB209Gsat	MMRRC	MGI: 3844824
Mouse: Tcf21-creERT2	Acharya et al., 2011	N/A
Oligonucleotides		
Primers for qPCR, see Table S7	This paper	N/A

Drop-seq primers, see Table S7	Macosko et al., 2015	N/A
Drop-seq beads	ChemGenes	Macosko201110
smFISH probes, see Table 7	LGC Biosearch Technologies	N/A
smHCR probes, see Table 7	Molecular Technologies	N/A
Software and Algorithms		
Drop-seq_tools (v1.12)	Macosko et al., 2015	http://mccarrolllab.com/dropseq/
Picard Tools (v2.6.0)	Broad Institute, 2016	http://broadinstitute.github.io/picard/
Samtools (v1.2)	Li et al., 2009	http://samtools.sourceforge.net/
STAR (v2.5.2b)	Dobin et al., 2013	https://github.com/alexdobin/STAR
R (v3.3.3)	R Core Team, 2017	https://www.R-project.org/
Seurat (v1.4.0.3)	Satija et al., 2015	https://github.com/satijalab/seurat
Seriation (v1.2–2)	Hahsler et al., 2008	https://CRAN.R-project.org/package=seriation
Monocle 2	Qiu et al., 2017	https://github.com/cole-trapnell-lab/monocle-release
Waterfall	Shin et al., 2015	https://omictools.com/waterfall-tool
som (v0.3.5.1)	Yan et al., 2016	https://CRAN.R-project.org/package=som
MATLAB R2017a	The MathWorks	https://www.mathworks.com/products/
Other		
Resource website for the publication	This paper	https://github.com/qianqianshao/Drop-seq_ST